Co-summability

From Linear to Non-linear Co-integration

by Vanessa Berenguer-Rico

Dissertation Presented in

Candidacy for the Degree of

Doctor of Philosophy

Dissertation Director: Jesús Gonzalo

Universidad Carlos III de Madrid

July 12, 2011

Co-summability

From Linear to Non-linear Co-integration

Por Vanessa Berenguer-Rico

Director: Jesús Gonzalo

Firma del Tribunal Calificador:

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Calificación:

Getafe, a de de 2011

To all those who educated me

Acknowledgments

En primer lugar, y, sí, por ello el más importante, me gustaría darle las gracias a Jesús Gonzalo. Intentaré sintetizar... Por haber dirigido esta tesis con tanta pedagogía, dedicación e inteligencia emocional y académica. Por los recursos invertidos, que nunca fueron escasos. Por sumar y co-sumar conmigo semana tras semana –cada viernes suponía un influjo de conocimiento para seguir avanzando. Por contribuir a hacer de la realización de esta tesis uno de mis hobbies y enseñarme, sin decírmelo, que los contratiempos presentes pueden llegar a convertirse en satisfacción futura. Por dejarme aprender tu método docente. Por lo motivante y divertido que es investigar contigo. Y gracias también por haber confiado en mí desde el primer día en aquel *Frontiers in Time Series Analysis* y haberme tenido ahí, en la frontera, durante todos estos años, enseñándome a caminar por ella. I will miss our econometric fridays.

While preparing this thesis, especially Chapter 1, I had the great opportunity to exchange ideas and thoughts on summability and co-summability with many great researchers. Some of them were visiting Madrid, others were at international conferences and workshops. I am indebted to J.J. Dolado, C.W.J. Granger, S. Johansen, O. Linton, P.C.B. Phillips, D. Politis, P. Robinson, M.H. Seo, W. Stute and C. Velasco for their valuable and constructive comments and suggestions that helped to improve upon previous versions of Chapter 1. I am especially grateful to all researchers that discussed this thesis during my job market seminars. I also want to thank participants at the 18th Annual Symposium of the SNDE, the 10th World Congress of the Econometric Society, the 2010 NBER-NSF Time Series Conference, 2010 European Winter Meetings of the Econometric Society, XXXV Simposio de Análisis Económico, the 4th Computational and Financial Econometrics, and the I_t Workshop in Time Series Econometrics for their helpful insights. Financial support from SEJ-2007-63098, CONSOLIDER 2010 (CSD 2006-00016) and EXCELECON S-2007/HUM-044 grants is gratefully acknowledged.

Al Departamento de Economía de la Universidad Carlos III de Madrid, gracias por todo el conocimiento transmitido y por los recursos humanos y materiales que siempre tuve a mi disposición. Gracias a todos mis profesores del Master y en especial a aquellos de los que tanta econometría he aprendido. Estoy en deuda con Carlos Velasco y Juanjo Dolado por sus sugerencias y comentarios, siempre constructivos, que han mejorado sustancialmente versiones previas de esta tesis. Gracias también a mis compañeros de doctorado, en especial a los de mi clase –Dolo, Eva, Joaquín Coleff, Joaquín Gomez, Lucila y Valery– que han amenizado las horas en Getafe.

Decir gracias es poco para lo que mis compañeros de viaje en el despacho 9.0.38 han significado para mí. Joaquín, sin tus innumerables ánimos del primer año de doctorado, posiblemente, ahora no estaría escribiendo estas líneas. Visto en perspectiva, no hay capítulo de esta historia en la que no hayas estado literalmente a mi lado. Gracias por aguantarme todo este tiempo estoicamente y por tu paciencia sin límite. Gracias también por introducir siempre tu argumento de economista-economista en nuestros interminables debates y por toda la economía que me has enseñado. Gracias, en mayúsculas, por ser mi mejor amigo en Madrid. Tengo que agradecerle a Heiko el haber subido al 9.0.38 y traer consigo una inconmensurable disponibilidad para ayudar tanto a nivel académico como personal. Gracias por los momentos que hemos compartido en la universidad y fuera de ella. Espero seguir compartiendo muchos momentos más.

A Noelia, gracias por haber venido a la Carlos III. Ahora se que tú eras lo que le faltaba a mi doctorado antes de que vinieras.

Y también gracias al Personal de Administración y Servicios de la Universidad. En especial, a Irene por darme el mejor soporte informático posible pero sobre todo por nuestras inolvidables vivencias; y también gracias a Amado, Elena y Rosi por amenizar todas y cada una de las mañanas del doctorado.

A mis compañeros de piso en Getafe y Madrid, gracias por ofrecer el ambiente necesario para la realización de una tesis doctoral y por los ratos de ocio que he vivido con todos vosotros.

Si bien, oficialmente, esta tesis comenzó en la Universidad Carlos III de Madrid, su gestación tuvo lugar en la Universitat de Barcelona. Me gustaría agraceder a todos mis profesores de econometría allí por cada una de sus clases, que despertaron mi, hasta la fecha, insaciable curiosidad por la econometría; y también por enseñarme a estudiarla desde esa doble perspectiva teórico-aplicada. Quique, he intentado no olvidar las cinco primeras letras de la palabra econometría; espero haberlo conseguido. Josep Lluís per a tu em falten paraules d'agraïment per poder expressar la meva gratitud. Tot el que em vas ensenyar m'ha sigut d'utilitat en algun moment del doctorat a la Carlos III. Gràcies per transmetrem la teva passió per les series temporals. I gràcies també per haver estat i per seguir estan sempre.

Val a dir que no tota l'economia que vaig aprendre a la UB me la van ensenyar els professors. Una gran part l'he après dels meus companys de classe, tant de llicenciatura com del doctorat. Gràcies Jordi per fer-me veure la economia amb ulls d'economista. Y gràcies també per totes les estones acadèmiques i no acadèmiques que em compartit i compartirem. A la Elena perquè treure una llicenciatura tenint-la a ella al costat es molt, però que molt, mes fàcil. Al David, al Dani, a la Rous i, de nou, al Jordi tinc que agrair els peer effects dels que vaig poder gaudir i aprofitar-me durant les classes del doctorat. Y a la Eli, por haber compartido conmigo, en el mítico despacho 108, uno de mis mejores años de estudiante y significar, a partir de entonces, alguien muy especial para mi.

A mis compañeras de piso en Barcelona que también ayudaron en la gestación de esta tesis. En especial, moltes gràcies a la Irene, la Cris i la Rosi per fer compatible l'aventura d'independitzar-se amb aprendre economia i, a mes a mes, fer-ho tant be i ensenyant-me tantes coses. Y moltes gràcies a la Rous, per haver compartit amb mi un any inoblidable i escoltar tot el que passava al temple economètric; ets asimptoticament genial.

Dicen que lo bueno siempre se deja para el final. Así pues, finalmente, quiero darles las gracias a mis padres, Carlos y Maribel, a mi hermano Carlos, a Elena y a Martina por todo el apoyo incondicional que he recibido durante la realización de esta tesis y a lo largo de mi vida; tanto más, cuanto más difícil era el momento que yo atravesaba y siempre obviando sus propias circunstancias. Esto no sólo es de agradacer, esto es para quererlos hasta el infinito. También quiero darle las gracias a mis tíos, Tonet y Rosaura, por ser, desde siempre, ejemplo docente e investigador. Y a mis amigas y amigos de Fraga por, a pesar de mis largas ausencias, seguir estando siempre ahí.

A mis padres y a mi hermano por ayudarme a hacer este sueño realidad.

"I developed a love of science and the scientific method. I think this is why I studied econometrics; it is the place where theory meets reality. It is the place where data on the economy tests the validity of economic theory."

Robert F. Engle III

Table of Contents

Introduction

Co-integration theory is an ideal framework to study linear relationships among persistent economic time series. Nevertheless, the intrinsic linearity in the ideas of integration and co-integration makes them unsuitable to study non-linear relationships between persistent processes. This drawback hinders the empirical analysis of modern macroeconomics which often deals with asymmetric responses to policy interventions, multiplicity of equilibria, transition between regimes or even log-linearized equilibria.

This thesis offers a theoretical and empirical framework to go beyond the linear world proposing and developing the idea of *order of summability*. A stochastic process y_t will be said to be summable of order δ if $\sum_{t=1}^{n} (y_t - E[y_t]) / n^{1/2+\delta} = O_p(1)$. This is a simple, general and useful concept able to synthetically describe key properties –such as persistence– of stochastic processes without relying on a particular data generating structure. To make this concept empirically relevant, in this thesis, an estimate of δ is obtained using a least squares regression. In addition, the asymptotic distribution of the estimator is derived. Due to the fact that, in general, this asymptotic distribution cannot be tabulated, subsampling methods are used to perform inferences. This machinery is applied to an extended version of the Nelson and Plosser database obtaining economic and econometric meaningful results.

The idea of order of summability and its associated econometric tools make it possible to study non-linear long run relationships under exactly the same logic as the one of co-integration theory. They have easily allowed (i) to define *balancedness* of a postulated relationship –a necessary condition for a correct specification– and (ii) to define non-linear long run relationships by means of the concept of *co-summability* –a direct extension of co-integration valid for non-linear equilibria. These two pieces will be relevant for both econometricians and economic theorists: for the former when specifying, estimating, and testing econometric models; for the latter when choosing functional forms to construct their theories.

The existing econometric literature assumes that a long run relationship indeed holds when estimating non-linear regression models. Nonetheless, non-linear transformations of persistent processes enlarge the possibilities for unbalancedness; and spurious regression problems can arise in non-linear models as well. To discriminate between these scenarios, in this thesis, tests for balancedness and co-summability are designed. The test for balancedness is based on the difference of the order of summability estimators of the time series in the model. The test for co-summability is based on the order of summability of the regression residuals. Both tests contribute to improve upon the statistical treatment of regression models involving non-linear transformations of persistent macroeconomic variables making them fully operative.

In the thesis, the practical strength of co-summability theory is shown through three empirical applications. Specifically, asymmetric preferences of central bankers, the environmental Kuznets curve and a present value model with a threshold discount factor for asset pricing are studied through the lens of co-summability.

Summarizing, this thesis opens a new set of econometric possibilities in the study of non-linear long run relationships generalizing co-integration theory. This generalization is presented in three different, although self-contained, chapters. Chapter 1 presents the order of summability and its associated econometric tools. Chapter 2 develops the ideas of balancedness and co-summability as well as the econometric theory to use them in practice. Their applicability is shown with two empirical illustrations: asymmetric preferences of central bankers and the environmental Kuznets curve. Chapter 3 derives the threshold present value model for asset pricing and tests it using the techniques developed in the two previous chapters. The corresponding appendices, collecting all the proofs, are at the end of each chapter.

Chapter 1

Summability

Abstract: The order of integration is valid to characterize linear processes; but it is not appropriate for non-linear worlds. This thesis proposes the concept of summability (a re-scaled partial sum of the process being $O_p(1)$) to handle non-linearities. Specifically, Chapter 1 shows that this new concept, $S(\delta)$: (i) generalizes $I(\delta)$; (ii) measures the degree of persistence as well as of the evolution of the variance; (iii) controls the balancedness of non-linear relationships; (iv) opens the door to the concept of co-summability which represents a generalization of co-integration for non-linear processes. To make this concept empirically applicable, an estimator for δ and its asymptotic properties are provided. The finite sample performance of subsampling confidence intervals is analyzed via a Monte Carlo experiment. The chapter finishes with the estimation of the degree of summability of the macroeconomic variables in an extended version of the Nelson-Plosser database.

1.1 Introduction

No one doubts that the concepts of integration and cointegration have been and still are very useful in time series econometrics. The former by producing a single parameter that was able to summarize the long-memory properties of a given time series. The latter by linking the existence of common trends to long-run linear equilibrium relationships. Thanks, amongst others, to the work by Dickey and Fuller (1979), Nelson and Plosser (1982), Phillips (1986), Engle and Granger (1987) and Johansen (1991), these two concepts are easily handled theoretically as well as empirically.

In parallel, non-linear time series models from a stationary perspective were introduced in the literature –see Granger and Teräsvirta (1993), Franses and van Dijk (2000), Fan and Yao (2003), and Teräsvirta, Tjϕstheim and Granger (2011) for some overviews. The introduction of persistent variables into non-linear models –see Park and Phillips (1999, 2001), de Jong and Wang (2005) or Pötscher (2004) for the study of transformations of integrated processes– produced a natural query: Which is the order of integration of these non-linear transformations? Such a question does not have a clear answer since the existing definitions of integrability do not properly apply. Integration is a linear concept.

This lack of definition has at least two important worrying consequences. First, in univariate terms, it implies that an equivalent synthetic measure of the stochastic properties of the time series, such as the order of integration, is not available to characterize non-linear time series. This does not only affect econometricians, but also economic theorists who cannot neglect important properties of actual economic variables when choosing functional forms to construct their theories. Second, from a multivariate perspective, it becomes troublesome to determine whether a non-linear model is balanced or not. Unbalanced equations are related to the familiar problem of misspecification, which is greatly enhanced when managing non-linear functions of variables having a persistence property. In linear setups, the concept of integrability did a good job dealing with balanced/unbalanced relations. However, in non-linear frameworks, the nonexistence of a synoptic quantitative measure makes it difficult to check the balancedness of a postulated model.

Additionally, this implies that a definition for non-linear co-integration is difficult to be obtained from the usual concept of integrability. To clarify this point, suppose $y_t = f(x_t, \theta) + u_t$, where $x_t \sim I(1)$, $u_t \sim I(0)$. For $f(\cdot)$ non-linear, the order of integration of $f(x_t, \theta)$, and hence that of y_t, may not be properly defined implying that the standard concept of co-integration is difficult to be applied. In fact, the literature on non-linear cointegration –see Park and Phillips (2001), Karlsen, Myklebust and Tjϕstheim (2007), Wang and Phillips (2009)– undertakes the whole analysis assuming the existence of a long-run relationship; something that should be tested in practice.

It was already stated in Granger and Hallman (1991) that a generalization of linear co-integration to a non-linear setup goes through proper extensions of the linear concepts of $I(0)$ and $I(1)$. This has led some authors to introduce alternative definitions. For instance, Granger (1995) proposed the concepts of Extended and Short Memory in Mean. However, these concepts are neither easy to calculate nor general enough to handle some types of non-linear long run relationships. And, furthermore, a measure of the order of the Extended memory is not available. Dealing with threshold effects in co-integrating regressions, Gonzalo and Pitarakis (2006) faced these problems and proposed, in a very heuristic way, the concept of summability (a re-scaled partial sum of the process being $Op(1)$). However, they did not emphasize the avail of such an idea.

In this chapter, we define summability properly and show its usefulness and generality. Specifically, we put forward several relevant examples in which the order of integrability is difficult to be established, but the order of summability can be easily determined. Moreover, we show that integrated time series are particular cases of summable processes, in the sense that the order of summability is the same as the order of integration. Hence, summability is a generalization of integrability. Furthermore, summability does not only characterize some properties of univariate time series, but also allows to easily study the balancedness of a postulated relationship –linear or not. And maybe more important, non-linear long run equilibrium relationships between non-stationary time series can be properly defined. In particular, the concept of co-summability, which can be applied to extend co-integration to non-linear frameworks, is developed in Chapter 2.

To make this concept empirically operational, we propose a statistical procedure to estimate and carry out inferences on the order of summability of an observed time series. This makes useful the concept of summability not only in theory but also in practice. To estimate the order of summability, we use an estimator introduced by McElroy and Politis (2007) to analyze the rate of convergence of an statistic and is based on a simple least squares regression. The inference on the true order of summability is based on the subsampling methodology developed in Politis, Romano and Wolf (1999). It is shown, by simulations, that the subsampling machinery works reasonably well in finite samples given the generality of the approach. Finally, the proposed methodology is used to estimate the order of summability of the macroeconomic time series in an extended version of Nelson-Plosser database.

The chapter is organized as follows. In the next section, the problems of using the order of integration to characterize non-linear processes are highlighted. In section 1.3, our proposed solution based on summability is described and its simple applicability showed. Section 1.4 describes the statistical tools –estimation and inference– to empirically deal with summable processes. In Section 1.5, an empirical application shows how to determine the order of summability in practice. Finally, Section 1.6 is devoted to some concluding remarks. All proofs are collected in the Appendix at the end of the chapter.

A word on notation. We use the symbol "$\Longrightarrow$" to signify convergence in distribution and weak convergence indistinctly, "$\xrightarrow{p}$" to signify convergence in probability. Stochastic processes such as the standard Brownian motion $W(r)$ are defined on $[0,1]$. Finally, all limits given in this chapter are taken as the sample size $n \to \infty$.

1.2 Order of Integration and Non-linear Processes

1.2.1 Order of Integration

Definition 1 *: A time series y_t is called an integrated process of order d (in short, an $I(d)$ process) if the time series of dth order differences $\Delta^d y_t$ is $I(0)$.*

A natural question that arises after reading this definition is: and what is an $I(0)$ process? Attempts to give a formal description of $I(0)$ processes exist in the literature. Engle and Granger (1987) give the following characterization.

Engle and Granger (EG) Characterization: *If $y_t \sim I(0)$ with zero mean then (i) the variance of y_t is finite; (ii) an innovation has only a temporary effect on the value of y_t; (iii) the spectrum of y_t, $f(\omega)$, has the property $0 < f(0) < \infty$; (iv) the expected length of time between crossing of $x = 0$ is finite; (v) the autocorrelations, ρ_k, decrease steadily in magnitude for large enough k, so that their sum is finite.*

Other characterizations have been used as well. Granger (1995) and Johansen (1995) used autoregressive and moving average representations, respectively. Müller (2008) and Davidson (2009) –among others– define an $I(0)$ as a process that satisfy the functional central limit theorem (FCLT). These latter definitions share the same spirit of our summability definition in Section 1.3. Nevertheless, in all cases, differences must be taken to discover the order of integration and the intrinsic linearity of the difference operator makes it difficult, if not impossible, to characterize –among others– non-linear processes. Integration is a linear concept.

1.2.2 Examples

Example 1 *: Alpha Stable i.i.d. Distributed Processes*

Let y_t be *i.i.d.* from some distribution $F \in D(\alpha)$, where $D(\alpha)$ denotes the domain of attraction of an α-stable law with $\alpha \in (0,2]$. y_t is strictly stationary; however, its second moments may not exist. The fact that such a process is *i.i.d.* could incline to think that this process is $I(0)$. However, if second moments do not exist, EG Characterization does not apply. Characterizations based on the FCLT could not be used either since they assume a standard Brownian motion in the limit. Hence, it becomes troublesome to establish the order of integration of y_t.

Example 2 *: An i.i.d. plus a Random Variable*

Consider the following process

$$y_t = z + e_t, \tag{1.1}$$

where $z \sim N(0, \sigma_z^2)$ and $e_t \sim i.i.d.(0, \sigma_e^2)$ are independent of each other. This process has the following properties

(i) $E[y_t] = 0$

(ii) $V[y_t] = \sigma_z^2 + \sigma_e^2$

(iii) $\gamma_y(k) = Cov(y_t, y_{t-k}) = \sigma_z^2$ for all $k > 0$.

Since it is a strictly stationary process, one could think that it is $I(0)$. However, the autocovariance function is not absolutely summable and its spectrum does not satisfy the required condition in EG Characterization[1]. If y_t is not $I(0)$, to attach any other order of integration to this stochastic process is not obvious. It is controversial to say y_t is $I(1)$ since $\Delta y_t = \Delta e_t$ is generally understood as an $I(-1)$; and it becomes difficult to choose any other number using the above definition of order of integration.

Dealing with non-linear processes similar problems are faced.

Example 3 *: Product of i.i.d. and Random Walk*

Let

$$w_t = \pi_t \eta_t, \tag{1.2}$$

where $\eta_t \sim i.i.d.\,(0,1)$ and

$$\pi_t = \pi_{t-1} + \varepsilon_t, \tag{1.3}$$

with $\pi_0 = 0$ and $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$ independent of η_t. Some properties of w_t are

(i) $E[w_t] = 0$

(ii) $V[w_t] = \sigma_\varepsilon^2 t$

(iii) $\gamma_w(h) = E[w_t w_{t-h}] = 0$.

[1] The autocovariance of the process in this example can be expressed as

$$\gamma(h) = \int_{-\pi}^{\pi} e^{ih\lambda} \left[\frac{\sigma_z^2 + \sigma_e^2}{2\pi} + \frac{\sigma_z^2}{\pi} \sum_{h=1}^{\infty} \cos(\lambda h) \right] d\lambda.$$

Then, the spectral density is

$$f(\lambda) = \frac{\sigma_z^2 + \sigma_e^2}{2\pi} + \frac{\sigma_z^2}{\pi} \sum_{h=1}^{\infty} \cos(\lambda h),$$

which diverges for all λ.

It is not obvious to attach an order of integration to this process. On one hand, the uncorrelation property (iii) could incline to think that w_t is $I(0)$. However, an $I(0)$ cannot have a trend in the variance according to EG Characterization. On the other hand, this unbounded variance could induce to suspect that the process is $I(1)$. Nevertheless, its first difference

$$\Delta w_t = \pi_t \eta_t - \pi_{t-1}\eta_{t-1},$$

cannot be $I(0)$ since, again,

$$V[\Delta w_t] = E[(\pi_t \eta_t)^2] + E[(\pi_{t-1}\eta_{t-1})^2] - 2E[\pi_t \pi_{t-1}\eta_t \eta_{t-1}] = (2t-1)\sigma_\varepsilon^2.$$

This means that w_t cannot be $I(1)$. It cannot be $I(2)$ either, since the variance of the second difference is

$$V[\Delta^2 w_t] = E[(\pi_t \eta_t)^2] + 4E[(\pi_{t-1}\eta_{t-1})^2] + E[(\pi_{t-2}\eta_{t-2})^2] = 6(t-1)\sigma_\varepsilon^2.$$

In fact, this process can be considered to be $I(\infty)$, in the sense that, the variance of $\Delta^d w_t$ depends on t regardless of the value of d –see Yoon (2005).

As pointed out by Granger (1995), non-linear transformations of highly heterogeneous or volatile processes, although uncorrelated, can induce high correlations. This can be seen by analyzing

$$q_t = \pi_t \eta_t^2, \tag{1.4}$$

where π_t and η_t are defined as before. The only difference is that now the *i.i.d.* sequence, η_t^2, is always positive. However, in this case,

$$E[q_t] = E[\pi_t \eta_t^2] = 0,$$

$$V[q_t] = E[q_t^2] = E[\pi_t^2 \eta_t^4] = E[\pi_t^2]E[\eta_t^4] = t\sigma_\varepsilon^2 \mu_4,$$

and

$$\gamma_q(h) = E[q_t q_{t-h}] = E[\pi_t \pi_{t-h}\eta_t^2 \eta_{t-h}^2] = E[\pi_t \pi_{t-h}]E[\eta_t^2 \eta_{t-h}^2] = (t-h)\sigma_\varepsilon^2 \sigma_\eta^4,$$

where $\mu_4 = E[\eta_t^4]$. Now, both variance and covariance depend on time. Hence, it can be seen how non-linear transformations of highly heterogenous processes can have an important impact on its stochastic properties. This impact will be hardly contemplated by the order of integration.

Example 4 *: Square of a Random Walk*

Consider now the square of the random walk defined in equation (1.3),

$$\pi_t^2 = \pi_{t-1}^2 + 2\pi_{t-1}\varepsilon_t + \varepsilon_t^2. \tag{1.5}$$

To establish the order of integration of this process is again not an obvious task. Granger (1995) considers that π_t^2 can be seen as a random walk with drift, hence, one could think that π_t^2 is $I(1)$. However,

$$V[\pi_t^2 - \pi_{t-1}^2] = E[\varepsilon_t^4] + 4(t-1)\sigma_\varepsilon^4 - \sigma_\varepsilon^4.$$

Again EG Characterization cannot be applied to $\Delta\pi_t^2$ or $\Delta^d\pi_t^2$.

Example 5 *: Product of Indicator Function and Random Walk*

Let

$$h_t = 1(v_t \leq \gamma)\pi_t, \tag{1.6}$$

where v_t is $i.i.d.\,(0,1)$ independent of ε_t, $1(\cdot)$ is the indicator function, and π_t is the random walk defined in (1.3). The variance and autocovariances of h_t depend on time, hence, one would think that it is $I(1)$. However, again, the variance of the first difference

$$\begin{aligned} V[\Delta h_t] &= E\left[(1(v_t \leq \gamma)\pi_t - 1(v_{t-1} \leq \gamma)\pi_{t-1})^2\right] \\ &= pt\sigma_\varepsilon^2 + p\,(t-1)\,\sigma_\varepsilon^2 - 2p^2\,(t-1)\,\sigma_\varepsilon^2 \\ &= [2p(1-p)\sigma_\varepsilon^2]t + p(2p-1)\sigma_\varepsilon^2, \end{aligned}$$

where $p = \Pr(v_t \leq \gamma)$. In fact, it can be considered, once again, that $h_t \sim I(\infty)$.

Example 6 *: Park and Phillips (1999, 2001)*

Similar incongruities to those encountered in previous examples appear when dealing with the non-linear transformations of $I(1)$ processes studied in Park and Phillips (1999, 2001); for instance, $e^{-\pi_t^2}$, $1/(1+\pi_t^2)$, $\log(|\pi_t|)$, or $(1+e^{-\pi_t})^{-1}$.

Example 7 *: Stochastic Unit Root and Explosive Processes*

Consider, on one hand, a stochastic unit root process

$$y_t = \rho_t y_{t-1} + \varepsilon_t, \tag{1.7}$$

where $y_0 = 0$ and $\rho_t \sim i.i.d.(\rho, \omega^2)$ is independent of $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$. On the other hand, contemplate the following explosive process

$$z_t = \phi z_{t-1} + \xi_t, \tag{1.8}$$

with $z_0 = 0$, $\phi > 1$ and $\xi_t \sim i.i.d.(0, \sigma_\xi^2)$. As in previous examples, to determine the order of integration of y_t and z_t is troublesome.

In all these examples the order of integrability is difficult to be calculated. The standard $I(d)$ classification is not sufficient to handle many stochastic processes.

1.3 A Solution Based on Summability

1.3.1 Order of Summability

The idea of order of summability of a stochastic process was initially introduced in a heuristic way in Gonzalo and Pitarakis (2006) when dealing with threshold effects in co-integrating regressions. In this section, the concept of summability is formalized and its generality, usefulness, and simplicity are asserted.

Definition 2 *: A stochastic process y_t with positive variance is said to be summable of order δ, represented as $S(\delta)$, if*

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1) \qquad \text{as } n \to \infty, \tag{1.9}$$

where δ is the minimum real number that makes S_n bounded in probability, m_t is a deterministic sequence, and $L(n)$ is a slowly-varying function[2]*.*

Note that, when possible, the order of summability will be determined by some Central Limit result. In the standard Central Limit Theorem –CLT–, for instance, $\delta = 0$ and $L(n)$ is just a constant. When the time series is a random walk, by the Functional Central Limit Theorem –FCLT– and the Continuous Mapping Theorem –CMT–, $\delta = 1$ and $L(n)$ is again a constant term. Although, in many circumstances $L(n)$ will be constant, in some situations[3] the asymptotic theory will enforce us to use an L function varying with n but slowly in the Karatama's sense.

From a more general perspective, the relationship between integrability and summability is discussed in the following two propositions.

Assumption 1: *Let y_t be the $I(d)$ process $\Delta^d y_t = C(L) u_t$*, where $u_t = \varepsilon_t 1(t > 0)$. ε_t has zero mean, is *i.i.d.*, and $E|\varepsilon_t|^r < \infty$ for $r \geq \max[4, -8d_0/(1+2d_0)]$ with $d_0 \in (-1/2, 1/2]$. *In addition,* $C(L) = \sum_{j=0}^{\infty} c_j L^j$, with $0 < |C(1)| < \infty$, $\sum_{j=0}^{\infty} c_j^2 < \infty$, and $\sum_{j=1}^{\infty} j^2 c_j^2 < \infty$.

Proposition 1 *: Under Assumption 1 if the time series y_t is $I(d)$ with $d \geq 0$, then it is $S(d)$.*

[2] A positive, Lebesgue measurable function L, on $(0, \infty)$ is slowly varying –in the Karatama's sense– at ∞ if

$$\frac{L(\lambda n)}{L(n)} \to 1 \quad (n \to \infty)\ \forall \lambda > 0.$$

(See Embrechts, Klüppelberg and Mikosh, 1999, p.564).

[3] Consider the case where the process y_t has density $f(x) = 1/|x|^3$ for $|x| > 1$. In that case, it is known (e.g., Romano and Siegel, 1986, Example 5.47) that

$$\frac{1}{[n \log n]^{1/2}} \sum_{t=1}^{n} y_t \Longrightarrow N(0,1).$$

Then, $L(n) = (1/\log n)^{1/2}$.

Next proposition deals with processes with negative orders of integration.

Proposition 2 : *Under Assumption 1 if the time series y_t is $I(-d)$ with $d = 1, 2, ... < \infty$, then it is $S(-0.5)$.*

Since negative integer orders of integration are not very relevant, only $d \geq 0$ will be considered. Hence, $I(d)$ processes are $S(d)$.

1.3.2 Examples

For all processes considered in Examples 1-7 the order of integration was not possible to be established. Next, for these examples, it is shown that the order of summability can be easily obtained.

Summability in Example 1 (*α-stable i.i.d. process*): Let y_t be symmetric around zero. By the Generalized Central Limit Theorem

$$S_n = \frac{1}{n^{\frac{1}{\alpha}}} L(n) \sum_{t=1}^{n} y_t \Longrightarrow S_\alpha,$$

where $S_\alpha \sim F \in D(\alpha)$. Hence, in this case the time series is said to be summable of order $\delta = (2-\alpha)/2\alpha$. For instance, a Cauchy distributed process ($\alpha = 1$) is $S(0.5)$.

Summability in Example 2 (*An i.i.d. plus a random variable*): From (1.1)

$$S_n = \frac{1}{n} \sum_{t=1}^{n} y_t = \frac{1}{n} \sum_{t=1}^{n} (z + e_t) = z + \frac{1}{n} \sum_{t=1}^{n} e_t \Longrightarrow z.$$

Therefore, y_t is $S(0.5)$.

Summability in Example 3 (*Product of i.i.d. and random walk*): It can be shown –see for instance, Park and Phillips (1988)– that

$$S_n = \frac{1}{\sigma_\varepsilon n} \sum_{t=1}^{n} \pi_t \eta_t \Longrightarrow \int_0^1 W_1(r) dW_2(r).$$

This means that $\pi_t \eta_t$ is $S(0.5)$ with, for instance, $L(n) = 1/\sigma_\varepsilon$.

For $\pi_t \eta_t^2$ note that,

$$Var \left[\sum_{t=1}^{n} \pi_t \eta_t^2 \right] = O(n^3).$$

Then, by the Chebyshev's inequality,

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} \pi_t \eta_t^2 = O_p(1),$$

which implies that $\pi_t \eta_t^2$ is $S(1)$.

These two cases show that summability takes into account persistence as well as the variance behavior through time.

Summability in Example 4 (*Squared of a random walk*): It is well known that

$$S_n = \frac{1}{n^2\sigma_\varepsilon^2}\sum_{t=1}^{n}\pi_t^2 \Longrightarrow \int_0^1 W^2(r)dr.$$

Hence, π_t^2 is $S(1.5)$ with, for instance, $L(n) = 1/\sigma_\varepsilon^2$.

Summability in Example 5 (*Product of indicator function and random walk*): In this case,

$$S_n = \frac{1}{n^{\frac{3}{2}}p\sigma_\varepsilon}\sum_{t=1}^{n}1(v_t \leq \gamma)\pi_t \Longrightarrow \int_0^1 W(r)dr,$$

implying that $1(v_t \leq \gamma)\pi_t$ is $S(1)$ with, for instance, $L(n) = 1/p\sigma_\varepsilon$.

Summability in Example 6 (*Park and Phillips, 1999 and 2001*): The order of summability of the processes considered in this example can be obtained by using the asymptotic theory developed in Park and Phillips (1999). Specifically, it can be shown that $e^{-\pi_t^2} \sim S(0)$, $1/(1+\pi_t^2) \sim S(0)$, $\log(|\pi_t|) \sim S(0.5)$, and $(1+e^{-\pi_t})^{-1} \sim S(0.5)$.

Summability in Example 7 (*STUR and Explosive processes*): Consider the STUR process defined in (1.7). For simplicity, let $\rho_t \sim i.i.d.(1,1)$, i.e. set $\rho = \omega^2 = 1$. From Leybourne, McCabe and Tremayne (1996), it can be shown that

$$S_n = \frac{1}{2^{n/2}}\sum_{t=1}^{n} y_t = O_p\left(1\right).$$

With respect the explosive process (1.8), from White (1958)

$$S_n = \frac{1}{\phi^n}\sum_{t=1}^{n} z_t = O_p\left(1\right).$$

Strictly speaking, the order of summability of y_t and z_t will be ∞. These are cases of non-summable processes.

As it can be seen with these examples, when possible, the order of summability is determined by some Central Limit result. When powers and products of integrated processes are considered, the order of summability can be easily obtained from the asymptotic theory for $I\left(d\right)$ processes. Nevertheless, other non-linear transformation, as those in Example 6, require new asymptotic theory results –see Park and Phillips (1999, 2001), Pötscher (2004) or de Jong and Wang (2005). Specifically, Park and Phillips (1999) showed that for all integrable functions $f\left(\pi_t\right)$, such as $f\left(\pi_t\right) = e^{-\pi_t^2}$ or $f\left(\pi_t\right) = 1/(1+\pi_t^2)$, $f\left(\pi_t\right) \sim S\left(0\right)$. However, the order of summability of asymptotically homogeneous tranformations[4], such as $h\left(\pi_t\right) = \log(|\pi_t|)$, $h\left(\pi_t\right) = (1+e^{-\pi_t})^{-1}$, or $h\left(\pi_t\right) = \pi_t^k$, is not

[4] A tranformation T is said to be asymptotically homogeneous iff $T\left(\lambda x\right) = \nu\left(\lambda\right)H\left(x\right) + R\left(x,\lambda\right)$ where $H\left(x\right)$ is locally integrable and $R\left(x,\lambda\right)$ is such that either (a) or (b) holds:

(a) $\left|R\left(x,\lambda\right)\right| \leq a\left(\lambda\right)P\left(x\right)$, where $\limsup_{\lambda\to\infty} a\left(\lambda\right)/\nu\left(\lambda\right) = 0$ and P is locally integrable.

(b) $\left|R\left(x,\lambda\right)\right| \leq b\left(\lambda\right)Q\left(\lambda x\right)$, where $\limsup_{\lambda\to\infty} b\left(\lambda\right)/\nu\left(\lambda\right) = 0$ and Q is locally integrable and vanishes at infinity, i.e. $Q\left(x\right) \to 0$ as $|x| \to \infty$.

The reader is refered to Park and Phillips (1999, 2001) for further technical details.

always the same. In fact, in these cases

$$\frac{1}{n\nu\left(\sqrt{n}\right)}\sum_{t=1}^{n}h\left(\pi_t\right)=O_p\left(1\right),$$

where $\nu\left(\sqrt{n}\right)$ is the homogeneous component of the function h. In particular, for $h\left(\pi_t\right)=\log(|\pi_t|)$, $\nu\left(\sqrt{n}\right)=\log\left(\sqrt{n}\right)$, for $h\left(\pi_t\right)=(1+e^{-\pi_t})^{-1}$, $\nu\left(\sqrt{n}\right)=1$, and for $h\left(\pi_t\right)=\pi_t^k$, $\nu\left(\sqrt{n}\right)=n^{k/2}$. Therefore, while $\log(|\pi_t|)\sim S(0.5)$ and $(1+e^{-\pi_t})^{-1}\sim S(0.5)$, $h\left(\pi_t\right)=\pi_t^k\sim S((k+1)/2)$.

1.3.3 Some Uses of Summability

In the same way integration constitutes the first step to check the balancedness of a linear relationship and to analyze cointegration, summability can be used to study non-linear long run relationships.

Definition 3 *: A postulated relationship*

$$y_t=f\left(x_t,\theta\right),$$

will be said to be balanced if $y_t\sim S\left(\delta_y\right)$, $z_t=f\left(x_t,\theta\right)\sim S\left(\delta_z\right)$*, and* $\delta_y=\delta_z$.

Once the balancedness of a non-linear model is established, the analysis of non-linear long run relationships can be done using the concept of co-summability.

Definition 4 *: Two summable stochastic processes,* $y_t\sim S\left(\delta_y\right)$ *and* $x_t\sim S\left(\delta_x\right)$*, will be said to be co-summable if there exists* $z_t=f\left(x_t,\theta\right)\sim S\left(\delta_y\right)$ *such that* $u_t=y_t-f(x_t,\theta)$ *is* $S(\delta_u)$*, with* $\delta_u=\delta_y-\delta$ *and* $\delta>0$*. In short,* $(y_t,z_t)\sim CS(\delta_y,\delta)$.

Co-summable processes will share an equilibrium relationship in the long run, i.e. an attractor $y_t=f(x_t,\theta)$ that can be linear or not. This type of equilibrium relationships will be usually established by the economic theory and have interesting econometric applications that include, for instance, transition behavior between regimes, multiplicity of equilibria, or non-linear responses to intervention policies. Applied researchers will be interested on estimating and testing these equilibria. A full treatment of co-summability in a regression framework is developed in Chapter 2.

1.4 Summability in Practice: Estimation and Inference

Following the same logic as in the integrated world, before any multivariate analysis –balancedness and co-summability–, it is necessary to develop the estimation and inference tools for the order of summability, δ, of univariate processes.

1.4.1 Estimation of δ

In this section, for simplicity reasons, it will be assumed $L(n) = 1$ in Definition 2. Therefore, the summability condition (1.9) becomes

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} (y_t - m_t) = O_p(1). \tag{1.10}$$

In addition, the next assumption is needed to implement our proposed estimation method of δ.

Assumption 2. $P(S_n = 0) = 0$ for all $n = 1, 2, 3, \ldots$

Our proposal to estimate δ, which follows the convergence rate estimator in McElroy and Politis (2007), is based on the transformation $U_n = \log S_n^2$. Hence, Assumption 2 is needed to avoid taking the logarithm of zero. Specifically, under Assumption 2, for a stochastic process y_t satisfying equation (1.10),

$$U_n = \log S_n^2 = \log \left(n^{-(1+2\delta)} \left(\sum_{t=1}^{n} (y_t - m_t) \right)^2 \right) = O_p(1). \tag{1.11}$$

Equation (1.11) can be written in regression model form as follows

$$Y_k = \beta \log k + U_k, \qquad k = 1, 2, \ldots, n, \tag{1.12}$$

where $\beta = 1 + 2\delta$, $Y_k = \log \left(\sum_{t=1}^{k} (y_t - m_t) \right)^2$, and $U_k = O_p(1)$.

We propose to estimate β by

$$\hat{\beta} = \frac{\sum_{k=1}^{n} Y_k \log k}{\sum_{k=1}^{n} \log^2 k}. \tag{1.13}$$

Given that $\beta = 1 + 2\delta$, the OLS estimator of δ is

$$\hat{\delta} = \frac{\hat{\beta} - 1}{2}.$$

Interestingly, the use of log-log regressions to estimate the degree of long memory of a time series has been used in other contexts as well. The original R/S statistic proposed by Hurst (1951) was driven by a log-log regression –see Beran (1994) or Giraitis et al. (1999) and the references therein. Besides, the well-known Geweke and Porter-Hudak (1983) and subsequent developments also entails a log-regression based on the periodogram ordinates at frequences near zero. The McElory and Politis (2007) procedure can be used to estimate, in the time domain, the memory parameter. Nevertheless, since it is not derived from a particular data generating structure, it can be used in a more general framework as the one considered in this chapter to estimate the order of summability of a stochastic process.

1.4.2 Asymptotic Properties

From (1.12) and (1.13)

$$\hat{\beta} - \beta = \frac{\sum_{k=1}^{n} U_k \log k}{\sum_{k=1}^{n} \log^2 k}. \tag{1.14}$$

Proposition 3 *(McElroy and Politis, 2007): Under Assumption 2,* $\hat{\beta} - \beta = o_p(1)$.

Remark: McElroy and Politis (2007) show that $\hat{\beta}$ is consistent under minimal assumptions. In our context, these assumptions are satisfied by definition of summable processes. Nonetheless, to the best of our knowledge, an asymptotic distribution for $\hat{\beta}$ has not yet been derived. The following proposition addresses this issue.

Proposition 4 *: Let* $x_t = y_t - m_t$. *Under Assumption 2, if*

$$S_n(r,\delta) = \frac{1}{n^{1/2+\delta}} \sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r,\delta), \tag{1.15}$$

where $D_x(r,\delta)$ *is some random process with positive variance, then*

$$\log n(\hat{\beta} - \beta) \Longrightarrow \int_0^1 U_x(r,\delta)\, dr, \tag{1.16}$$

with $U_x(r,\delta) = \log\left[\left(r^{-1/2-\delta} D_x(r,\delta)\right)^2\right]$.

Remark: When x_t is $i.i.d.(0,1)$, by the FCLT

$$S_n(r,0) = \frac{1}{n^{1/2}} \sum_{t=1}^{[nr]} x_t \Longrightarrow W(r).$$

Therefore, (1.16) becomes

$$\log n(\hat{\beta} - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-1/2} W(r)\right)^2\right] dr.$$

Similarly, if x_t is a standard random walk, then

$$S_n(r,1) = \frac{1}{n^{3/2}} \sum_{t=1}^{[nr]} x_t \Longrightarrow \int_0^r W(r) dr,$$

and

$$\log n(\hat{\beta} - \beta) \Longrightarrow \int_0^1 \log\left[\left(r^{-3/2} \int_0^r W(r) dr\right)^2\right] dr.$$

Remark: In many cases, $L(n) \neq 1$ but still $L(n) = c$, a constant different from zero. In such a case, regression (1.12) becomes

$$Y_k = \alpha + \beta \log k + U_k, \tag{1.17}$$

with $\alpha = -2\log c$. Notice that any c satisfies Definition 2. Therefore, α is not identified. Nevertheless, it is straightforward to get rid of it by substracting the first observation in regression (1.17) and estimating the model

$$Y_k^* = \beta \log k + U_k^*, \tag{1.18}$$

where $Y_k^* = Y_k - Y_1$ and $U_k^* = U_k - U_1$. The modified OLS estimator

$$\hat{\beta}^* = \frac{\sum_{k=1}^n Y_k^* \log k}{\sum_{k=1}^n \log^2 k},$$

satisfies the same asymptotic properties than those of $\hat{\beta}$.

An alternative way to take into account α could be using

$$\tilde{\beta} = \frac{\sum_{k=1}^n (Y_k - \bar{Y})(\log k - \overline{\log n})}{\sum_{k=1}^n (\log k - \overline{\log n})^2}. \tag{1.19}$$

In general, the lack of identification of α complicates the properties of $\tilde{\beta}$. For this reason, in this chapter only $\hat{\beta}^*$ is considered and consequently $\hat{\delta}^* = (\hat{\beta}^* - 1)/2$.

1.4.3 Subsampling Confidence Intervals

In general, the asymptotic distribution of $\hat{\beta}^*$ cannot be tabulated. Nevertheless, subsampling methods can be used to undertake inferences on the order of summability independently of its true value.

Subsampling is consistent under minimal assumptions. The most general result shown in Politis, Romano and Wolf (1999) requires that:

(i) the estimator, properly normalized, has a limiting distribution

(ii) the distribution functions of the normalized estimator based on the subsamples (of size b) have to be on average close to the distribution function of the normalized estimator based on the entire sample with $\log b/\log n \to 0$, $b/n \to 0$, $b \to \infty$

(iii) the sequence of the subsampling statistic $Z_{n,b,k} = \log b(\hat{\beta}^*_{n,b,k} - \beta)$, where $\hat{\beta}^*_{n,b,k}$ is the subsample estimator version of $\hat{\beta}^*$, has α-mixing coefficients, $\alpha_{n,b}(h)$, such that $n^{-1}\sum_{h=1}^n \alpha_{n,b}(h) \to 0$ as $n \to \infty$.

Conditions (i) and (ii) are guaranteed by Proposition 4. To show that the α-mixing condition (iii) holds in this context is beyond the scope of this thesis. The adequacy of the subsampling approach is analyzed via simulations using the twelve data generating processes –DGP– in Table 1.

Table 1: Data Generating Processes : $y_t = m_t + x_t$

$y_{1t} = m_t + \varepsilon_t,\ \varepsilon_t \sim iidN(0,1)$	$y_{7t} = m_t + \Delta^{0.3}\pi_t$
$y_{2t} = m_t + \pi_t,\ \pi_t = \sum_{j=1}^{t} \varepsilon_j$	$y_{8t} = m_t + z + \varepsilon_t,\ z \sim N(0,1) \perp \varepsilon_t$
$y_{3t} = m_t + \sum_{j=1}^{t} \pi_j$	$y_{9t} = m_t + \eta_t\pi_t,\ \eta_t \sim iidN(0,1) \perp \varepsilon_t$
$y_{4t} = m_t + \xi_t,\ \xi_t \sim iidCauchy$	$y_{10t} = m_t + \eta_t^2\pi_t,\ \eta_t \sim iidN(0,1) \perp \varepsilon_t$
$y_{5t} = m_t + \pi_t^2$	$y_{11t} = m_t + 1(v_t \leq 0)\pi_t,\ v_t \sim iidN(0,1) \perp \varepsilon_t$
$y_{6t} = m_t + t\varepsilon_t$	$y_{12t} = m_t + \log(\lvert\pi_t\rvert)$

Performance of subsampling is mainly measured by coverage probability, denoted CP, of two-sided nominal 95% symmetric intervals for δ. We also present the mean of the estimated $\delta's$ and the median lower and upper bounds of the estimated confidence intervals. These measures are denoted by $\bar{\delta}^*$, I_{low}, and I_{up}, respectively. The experiment is based on 1000 replicas and three different sample sizes $n = \{100, 200, 500\}$. Subsample size is $b = \sqrt{n}$. Results are collected in Table 2.

Table 2: Performance of subsampling intervals for δ. No Deterministic Components: $m_t = 0$

DGP	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n = 100$				$n = 200$				$n = 500$			
1 – $S(0)$	0.991	-0.004	-0.699	0.659	0.995	0.005	-0.607	0.566	0.991	0.000	-0.521	0.470
2 – $S(1)$	0.832	0.863	0.383	1.307	0.804	0.880	0.455	1.258	0.807	0.900	0.541	1.220
3 – $S(2)$	0.747	1.634	0.982	2.262	0.797	1.673	1.034	2.292	0.863	1.723	1.076	2.348
4 – $S(0.5)$	0.986	0.496	-0.414	1.387	0.992	0.521	-0.261	1.309	0.994	0.519	-0.185	1.187
5 – $S(1.5)$	0.905	1.516	0.701	2.192	0.900	1.519	0.771	2.107	0.904	1.510	0.828	2.049
6 – $S(1)$	0.990	0.862	-0.052	1.694	0.997	0.891	0.028	1.675	1.000	0.899	0.096	1.635
7 – $S(0.7)$	0.939	0.613	0.038	1.135	0.954	0.627	0.141	1.054	0.949	0.639	0.223	0.998
8 – $S(0.5)$	0.942	0.430	-0.213	1.007	0.929	0.401	-0.149	0.915	0.930	0.447	-0.024	0.875
9 – $S(0.5)$	0.988	0.507	-0.330	1.255	0.984	0.516	-0.206	1.164	0.983	0.501	-0.144	1.063
10 – $S(1)$	0.947	1.171	-0.106	2.311	0.952	1.167	0.099	2.127	0.954	1.124	0.220	1.894
11 – $S(1)$	0.598	0.689	0.220	1.104	0.644	0.743	0.325	1.140	0.650	0.767	0.389	1.105
12 – $S(0.5)$	0.844	0.557	0.041	0.977	0.801	0.630	0.196	0.988	0.705	0.694	0.353	0.982

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\bar{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b = \sqrt{n}$.

The performance of the subsampling method is adequate in general[5]. The coverage probability is around its nominal level and the mean estimated order of summability close to its true value. The subsampling confidence intervals, although wide, get narrower as the sample size increases. The amplitude of the intervals in small samples is basically a direct consequence of not assuming anything about the DGP of the analyzed time series.

1.4.4 Deterministic Components

Until now it has been assumed m_t to be known but this is not the case in practice. As in the integrated world, the presence of deterministic components can affect the estimation of the order of summability.

Let

[5]Notice that the coverage probability for cases 11 and 12 is very poor. Nonetheless, the consideration of deterministic components improve dramatically the coverage probability, as it can be seen in Tables 3 and 4.

$$y_t = m_t + x_t,$$

where
$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t \Longrightarrow D_x(\delta) \quad \text{and} \quad \frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} m_t \to \mu,$$
with $D_x(\delta) \equiv D_x(1,\delta)$ being a random variable with positive variance and μ a constant different from zero.

Consider the following two situations:

a. If $\delta > \gamma$, then
$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t + o(1) \Longrightarrow D_x(\delta).$$

b. If $\delta < \gamma$, then
$$\frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\gamma}}\sum_{t=1}^{n} m_t + o_p(1) \xrightarrow{p} \mu.$$

When $\delta < \gamma$, the order of the deterministic component dominates and it will be confused with the order of summability. Admittedly, even when $\delta > \gamma$, the deterministic components, if not properly considered, can affect the order of summability estimation in finite samples. Although not reported here, for space reasons, Monte Carlo experiments reveal the existence of an important bias effect when deterministic components are present and not properly taken into consideration. Therefore, in order to analyze the order of summability a proper technique to deal with these elements is needed.

Essentially, what is required is an estimator $\hat{m}_t$ such that
$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}(y_t - \hat{m}_t) \Longrightarrow D_x^*(\delta). \tag{1.20}$$
In other words, the order of summability of y_t is not affected by substracting $\hat{m}_t$.

Three usual parametric forms for m_t will be considered: $m_t = m_0$, $m_t = m_0 + m_1 t$, and $m_t = m_0 + m_1 t + m_2 t^2$. For these three cases, a proper treatment of the deterministic components is derived.

Constant Term Case: Let
$$y_t = m_0 + x_t,$$
where m_0 is a constant and $x_t \sim S(\delta)$ such that
$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n} x_t \Longrightarrow D_x(\delta).$$
Assume that only y_t is observed. The standard proposal of demeaning y_t by its arithmetic mean is problematic in this context because
$$\sum_{t=1}^{n}(y_t - \bar{y}) = 0. \tag{1.21}$$

Therefore, the true order of summability cannot be recovered. Next proposition shows that the partial mean

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j,$$

is an alternative operational choice in the sense of satisfying (1.20). Basically, the proposed $\hat{m}_t$ is the recursive version of the sample mean, $\bar{y}$.

Proposition 5 *: Consider the following DGP*

$$y_t = m_0 + x_t, \tag{1.22}$$

where m_0 is an unknown constant and

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r,\delta),$$

with $D_x(0,\delta) = 0$. If

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j, \tag{1.23}$$

then

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}(y_t - \hat{m}_t) \Longrightarrow D_x(1,\delta) - \int_0^1 r^{-1} D_x(r,\delta)dr.$$

Table 3 reports the performance of the subsampling confidence intervals after partially demeaning the processes described in Table 1 when $m_t = m_0 = 10$. Results do not depend on the value of m_0.

Table 3: Performance of subsampling intervals for δ. Constant Term: $m_t = 10$

DGP	CP	$\overline{\delta}^*$	I_{low}	I_{up}	CP	$\overline{\delta}^*$	I_{low}	I_{up}	CP	$\overline{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n = 100$				$n = 200$				$n = 500$			
$1 - S(0)$	0.982	0.085	-0.613	0.720	0.984	0.072	-0.523	0.618	0.987	0.061	-0.443	0.515
$2 - S(1)$	0.896	0.838	0.232	1.339	0.885	0.878	0.346	1.322	0.882	0.894	0.453	1.286
$3 - S(2)$	0.698	1.608	0.971	2.208	0.792	1.655	0.996	2.262	0.860	1.715	1.065	2.337
$4 - S(0.5)$	0.970	0.420	-0.424	1.185	0.969	0.443	-0.329	1.132	0.967	0.455	-0.171	1.039
$5 - S(1.5)$	0.752	1.208	0.378	1.956	0.788	1.266	0.506	1.957	0.814	1.305	0.624	1.920
$6 - S(1)$	0.981	0.775	-0.108	1.542	0.992	0.805	-0.020	1.555	0.999	0.822	0.049	1.515
$7 - S(0.7)$	0.970	0.582	-0.092	1.160	0.976	0.609	0.041	1.099	0.979	0.608	0.145	1.021
$8 - S(0.5)$	0.825	0.091	-0.594	0.736	0.707	0.071	-0.540	0.606	0.544	0.059	-0.442	0.524
$9 - S(0.5)$	0.985	0.398	-0.365	1.102	0.986	0.420	-0.259	1.041	0.986	0.443	-0.167	0.964
$10 - S(1)$	0.910	0.856	0.018	1.568	0.911	0.897	0.146	1.594	0.900	0.915	0.242	1.513
$11 - S(1)$	0.812	0.602	-0.134	1.291	0.831	0.667	0.008	1.278	0.841	0.711	0.123	1.271
$12 - S(0.5)$	0.943	0.525	-0.032	1.019	0.923	0.538	0.075	0.934	0.922	0.539	0.182	0.853

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\overline{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b = \sqrt{n}$.

Results are similar or even better than those obtained without deterministic components. For this reason, we recommend to always partially demean the processes.

Linear Trend Case: Let

$$y_t = m_0 + m_1 t + x_t,$$

where $x_t \sim S(\delta)$ in the sense that

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} x_t \Longrightarrow D_x(\delta),$$

as before. Next Proposition shows how to deal with the deterministic components in this case.

Proposition 6 *: Consider the following DGP*

$$y_t = m_0 + m_1 t + x_t, \tag{1.24}$$

where m_0 and m_1 are unknown parameters and

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r, \delta),$$

with $D_x(0,\delta)=0$. *If*

$$\hat{m}_t=\frac{1}{t}\sum_{j=1}^{t}y_j-\frac{2}{t}\sum_{j=1}^{t}\left(y_j-\frac{1}{j}\sum_{i=1}^{j}y_i\right), \tag{1.25}$$

then

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t)\Longrightarrow D_x(1,\delta)-3\int_0^1 r^{-1}D_x(r,\delta)\,dr.$$

Notice that in the linear trend case, the appropriate $\hat{m}_t$ consists, basically, in a double partial –or recursive– demeaning procedure[6]. Table 4 summarizes the performance of subsampling confidence intervals after properly detrending the DGPs in Table 1 when $m_t=m_0+m_1t=10+2t$. As in the previous case, results do not depend on the particular choices of m_0 and m_1.

Table 4: Performance of subsampling intervals for δ. Linear Trend: $m_t=10+2t$

DGP	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}	CP	$\bar{\delta}^*$	I_{low}	I_{up}
$S(\delta)$	$n=100$				$n=200$				$n=500$			
1 – $S(0)$	0.933	0.282	-0.428	0.927	0.949	0.264	-0.359	0.831	0.953	0.228	-0.292	0.703
2 – $S(1)$	0.918	0.817	0.176	1.380	0.907	0.834	0.271	1.327	0.900	0.872	0.391	1.289
3 – $S(2)$	0.788	1.581	0.811	2.285	0.854	1.637	0.889	2.328	0.931	1.705	0.989	2.363
4 – $S(0.5)$	0.958	0.504	-0.274	1.174	0.965	0.501	-0.194	1.106	0.956	0.499	-0.098	1.028
5 – $S(1.5)$	0.726	1.096	0.329	1.816	0.755	1.144	0.433	1.818	0.799	1.198	0.539	1.790
6 – $S(1)$	0.973	0.727	-0.151	1.477	0.982	0.750	-0.058	1.464	0.997	0.795	0.033	1.473
7 – $S(0.7)$	0.978	0.616	-0.057	1.214	0.986	0.613	0.032	1.123	0.989	0.642	0.152	1.052
8 – $S(0.5)$	0.928	0.283	-0.429	0.929	0.912	0.273	-0.336	0.846	0.814	0.233	-0.280	0.726
9 – $S(0.5)$	0.985	0.456	-0.312	1.131	0.988	0.451	-0.220	1.080	0.991	0.467	-0.141	1.023
10 – $S(1)$	0.849	0.748	-0.047	1.436	0.858	0.770	0.055	1.411	0.865	0.805	0.150	1.393
11 – $S(1)$	0.794	0.621	-0.113	1.279	0.803	0.654	-0.030	1.254	0.832	0.707	0.076	1.281
12 – $S(0.5)$	0.928	0.559	-0.008	1.065	0.929	0.554	0.093	0.972	0.900	0.574	0.209	0.885

CP denotes the coverage probability of two-sided nominal 95% symmetric intervals. $\bar{\delta}^*$ represents the mean of the estimated orders of summability. I_{low} and I_{up} are the median of the lower and upper bounds of the intervals, respectively. 1000 replicas are used. Subsample size is $b=\sqrt{n}$.

[6] Other proper detrending procedures work too. We thank Franco Peracchi for pointing out the alternative methodology of applying a recursive OLS detrending, i.e. $\hat{m}_t=\hat{\alpha}_t+\hat{\beta}_t t$ where $\hat{\alpha}_t=(1/t)\sum_{j=1}^{t}y_j-\hat{\beta}_t(1/t)\sum_{j=1}^{t}j$ and $\hat{\beta}_t=\sum_{j=1}^{t}\left(y_j-(1/t)\sum_{j=1}^{t}y_j\right)\left(j-(1/t)\sum_{j=1}^{t}j\right)/\sum_{j=1}^{t}\left(j-(1/t)\sum_{j=1}^{t}j\right)^2$. This choice will be particularly interesting when fractional deterministic trends are present.

Results in Table 4 show that the proposed detrending method $\hat{m}_t$ performs adequately in finite samples.

Quadratic Trend Case: Let

$$y_t = m_0 + m_1 t + m_2 t^2 + x_t,$$

where $x_t \sim S(\delta)$ such that

$$\frac{1}{n^{\frac{1}{2}+\delta}} \sum_{t=1}^{n} x_t \Longrightarrow D_x(\delta),$$

as before. The proposed $\hat{m}_t$ in this case is

$$\hat{m}_t = \frac{1}{t}\sum_{j=1}^{t} y_j - \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i\right) - \frac{3}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i - \frac{2}{j}\sum_{i=1}^{j}\left(y_i - \frac{1}{i}\sum_{h=1}^{i} y_h\right)\right).$$

Essentially, this transformation implies a triple partial demeaning procedure. It can be shown that the use of this $\hat{m}_t$ does not alter the order of summability of $y_t - \hat{m}_t$ and the finite sample performance is adequate (these results are available from the authors upon request).

Remark: It can be shown that if the order of the trend that is substracted is higher than the true one, then the order of summability of the detrended process, $y_t - \hat{m}_t$, is preserved; that is, it has the same order of summability that y_t. However, because of inefficiency issues, in general, it is not recommended to substract a very high polynomial trend.

Overall, the methodology proposed in this section to estimate the order of summability works reasonably well in finite samples. It is important to notice that our method does not assume any knowledge about the model generating the data. The trade off is that the confidence intervals are not very narrow.

1.5 Empirical Application

After Nelson and Plosser (1982) accounted for unit root behavior in almost all the fourteen U.S. macroeconomic time series in their database, many researchers have used the same dataset to confirm or refuse their conclusions with alternative approaches. In what follows, we contribute to this literature by applying the above developed methodology to estimate and infer the order of summability of the time series included in an extended version of the Nelson and Plosser (1982) database[7]. As a novelty, we do not impose any linearity assumption.

More precisely, we estimate the order of summability of the fourteen macroeconomic aggregates with $\hat{\delta}^* = (\hat{\beta}^* - 1)/2$ and derive the subsampling confidence intervals, denoted by (I_L^*, I_U^*). It is well known in the literature that deterministic components are an important issue for these time series.

[7] The data have been downloaded from P.C.B. Phillips' webpage.

Since the order of the deterministic trend is unknown, we propose to use in practice a traditional graphical device. If a trending behavior is observed, include *at least* a linear trend. If the time series evolve around a constant, consider *at least* a constant term. Using this device and knowing that it is always better to substract a higher than a lower order trend than the true one, a quadratic trend has been considered for all the variables but interest and unemployment rates. Results are shown in Table 5.

Table 5: Order of Summability. Estimation and Inference

log(variable)	Order of Summability		
quadratic trend	$\hat{\delta}^*$	I_L^*	I_U^*
consumer price index	2.369	1.112	3.625
employment	0.579	0.185	0.973
gnp deflator	0.900	0.168	1.631
nominal gnp	1.031	0.557	1.505
industrial production	0.738	0.082	1.393
gnp per capita	0.938	0.278	1.599
real gnp	0.898	0.287	1.510
wages	0.961	0.341	1.580
real wages	1.070	0.320	1.821
S&P	0.702	0.121	1.283
money	0.913	0.279	1.548
velocity	0.576	-0.010	1.163
linear trend	$\hat{\delta}^*$	I_L^*	I_U^*
interest	0.934	0.359	1.509
unemployment	0.162	-0.603	0.928

$\hat{\delta}^*$ denotes the estimated order of summability. I_L^* and I_U^* denote the lower and upper bounds of the corresponding subsampling intervals.

Observe that the variable with a lower order of summability is unemployment rate and the one with the highest the consumer price index. On the other hand, variables like nominal and real GNP, stock of money, wages, industrial production or S&P share similar orders of summability, around one. The amplitude of the confidence intervals is in line with the wide confidence intervals reported in Stock (1991) for the largest autoregressive root and in Arteche and Orbe (2005) for the fractional order of integration. Notice that our methodology does not assume any model for the data.

Overall, the estimated orders of summability of the fourteen macroeconomic variables seem to be quite reasonable in economic and econometric terms. Regarding the latter aspect of the empirical exercise, we would like to highlight the similarities of our results with those found in the fractional literature. With respect the economic content of the results, as already stated, variables like real and nominal GNP, industrial production, or nominal money have similar orders of summability and higher than those of unemployment or velocity of money. Additionally, in a heuristic way, it can be seen that these results do not go against the quantity theory of money.

1.6 Conclusion

Time Series Econometrics has not been able to properly handle non-linearities with persistent variables. This is mainly due to the fact that the concept of integration, and consequently cointegration, is too linear and not always well defined for non-linear processes. This lack of a proper definition has two important multivariate consequences. First, it is not possible to characterize the balancedness of a non-linear postulated model relating persistent variables. This is a necessary condition for an appropriate model specification. Second, co-integration cannot be directly extended to analyze non-linear long run relationships. The concept of summability is able to solve these problems. This chapter shows how to calculate, estimate, and undertake inference on the order of summability, δ.

1.7 Appendix

Proof of Proposition 1: Applying the Beveridge-Nelson decomposition as in Phillips and Solo (1992)

$$\Delta^d y_t = C(1) u_t + \tilde{u}_{t-1} - \tilde{u}_t,$$

with

$$\tilde{u}_t = \tilde{C}(L) u_t = \sum_{j=0}^{\infty} \tilde{c}_j L^j u_t = \sum_{j=0}^{\infty} \sum_{k=j+1}^{\infty} c_k u_{t-j}.$$

Now,

$$y_t = C(1) \Delta^{-d} u_t + \Delta^{-d} (\tilde{u}_{t-1} - \tilde{u}_t),$$

and

$$\frac{1}{n^{1/2+d}} \kappa(n,d)^{-1/2} \sum_{t=1}^{n} y_t = C(1) \Delta^{-d} \frac{1}{n^{1/2+d}} \kappa(n,d)^{-1/2} \sum_{t=1}^{n} u_t - \frac{1}{n^{1/2+d}} \kappa(n,d)^{-1/2} \Delta^{-d} \tilde{u}_n, \quad (1.26)$$

where

$$\kappa(n,d) = \begin{cases} \frac{\sigma_u^2 \Gamma(1-2d_0)}{(1+2d_0)\Gamma(1+d_0)\Gamma(1-d_0)} & if\ d > 1/2\ and\ d \neq \frac{2k+1}{2}\ \forall k \in \mathbb{N} \\ \frac{\sigma_u^2}{\pi} \log n & if\ d = \frac{2k+1}{2}\ \forall k \in \mathbb{N} \end{cases},$$

and $\Gamma(\cdot)$ denotes the gamma function.

Boundedness in probability of the first component of the right hand side of equation (1.26) was shown by Liu (1998). Hence, it remains to show boundedness in probability of the second term. To this end, without loss of generality, consider the case $d \in (0, 1/2)$ in which

$$\Delta^{-d} = \sum_{i=0}^{\infty} a_i L^i,$$

with $a_i = O\left(j^{d-1}\right)$. Note that

$$Var\left[\frac{1}{n^{1/2+d}}\Delta^{-d}\tilde{u}_n\right] = \frac{1}{n^{1+2d}}Var\left[\Delta^{-d}\tilde{u}_n\right] = \frac{1}{n^{1+2d}}Var\left[\sum_{i=0}^{\infty} a_i\tilde{u}_{n-i}\right] = \frac{1}{n^{1+2d}}\sum_{i=0}^{\infty} a_i^2 Var[\tilde{u}_{n-i}],$$

where

$$Var[\tilde{u}_{n-i}] = Var\left[\sum_{j=0}^{\infty}\tilde{c}_j u_{n-i-j}\right] = \sum_{j=0}^{\infty}\tilde{c}_j^2 Var\left[u_{n-i-j}\right] = \sigma_u^2\sum_{j=0}^{\infty}\tilde{c}_j^2.$$

Therefore,

$$Var\left[\frac{1}{n^{1/2+d}}\Delta^{-d}\tilde{u}_n\right] = \frac{\sigma_u^2}{n^{1+2d}}\sum_{i=0}^{\infty} a_i^2\sum_{j=0}^{\infty}\tilde{c}_j^2 = O\left(1\right),$$

implying

$$\frac{1}{n^{1/2+d}}\Delta^{-d}\tilde{u}_n = O_p\left(1\right).$$

Then $y_t \sim S\left(\delta\right)$. **Q.E.D.**

Proof of Proposition 2: The sum of y_t is

$$\sum_{t=1}^{n} y_t = C(1)\sum_{t=1}^{n}\Delta^d u_t - \Delta^d\tilde{u}_n = A_n - B_n,$$

where $A_n = C(1)\sum_{t=1}^{n}\Delta^d u_t$ and $B_n = \Delta^d\tilde{u}_n$. By definition of $\tilde{u}_t$,

$$B_n = \Delta^d\tilde{u}_n = O_p(1),$$

for all $d = 1, 2, ... < \infty$. With respect A_n note that,

$$C(1) < \infty,$$

and

$$\sum_{t=1}^{n}\Delta^d u_t = \Delta^{d-1}\sum_{t=1}^{n}\Delta u_t = \Delta^{d-1}u_n = O_p(1),$$

for all $d = 1, 2, ... < \infty$. Therefore,

$$A_n = C(1)\sum_{t=1}^{n}\Delta^d u_t = O_p(1),$$

as well. And, all together implies that

$$\sum_{t=1}^{n} y_t = A_n - B_n = O_p(1),$$

or equivalently that $y_t \sim S(-0.5)$. **Q.E.D.**

Proof of Proposition 3: By Assumption 2 and definition of summable process, U_k is $O_p(1)$. Hence, Theorem 3.1. in McElroy and Politis (2007) applies. **Q.E.D.**

Proof of Proposition 4: Expression (1.14) can be rewritten as

$$\log n\left(\hat{\beta} - \beta\right) = \frac{\frac{1}{n\log n}\sum_{k=1}^{n} U_k \log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n} \log^2 k}.$$

The denominator satisfies

$$\frac{1}{n\log^2 n}\sum_{k=1}^{n} \log^2 k \to 1 \qquad \text{as} \qquad n \to \infty.$$

With respect the numerator

$$\begin{aligned}
\frac{1}{n\log n}\sum_{k=1}^{n} U_k \log k &= \frac{1}{n\log n}\sum_{k=1}^{n} \log\left[\left(\frac{1}{k^{1/2+\delta}}\sum_{t=1}^{k} x_t\right)^2\right]\log k \\
&= \frac{1}{n\log n}\sum_{k=1}^{n} \log\left[\left(\frac{n^{1/2+\delta}}{k^{1/2+\delta}}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} x_k\right)^2\right]\log k \\
&= \frac{1}{n\log n}\sum_{k=1}^{n} \log\left[\left(\left(\frac{n}{k}\right)^{1/2+\delta}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} x_t\right)^2\right]\left(\log\left(\frac{k}{n}\right) + \log n\right) \\
&= \frac{1}{n\log n}\sum_{k=1}^{n} \left(\log\left[\left(\left(\frac{n}{k}\right)^{1/2+\delta}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} x_t\right)^2\right]\log\left(\frac{k}{n}\right)\right) \\
&\quad + \frac{1}{n}\sum_{k=1}^{n} \log\left[\left(\left(\frac{n}{k}\right)^{1/2+\delta}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} x_t\right)^2\right].
\end{aligned}$$

Let

$$U_{nk} = \log\left[\left(\left(\frac{n}{k}\right)^{1/2+\delta}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{k} x_t\right)^2\right],$$

and its D-space analog

$$U_n(r,\delta) = \log\left[\left(r^{-1/2-\delta}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{[nr]} x_t\right)^2\right],$$

which

$$U_n(r,\delta) \Longrightarrow \log\left[\left(r^{-1/2-\delta} D_x(r,\delta)\right)^2\right].$$

Now consider,

$$\begin{aligned} \frac{1}{n}\sum_{k=1}^{n} U_{nk}\log\left(\frac{k}{n}\right) &= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left[\log\left(\frac{k}{n}\right)+\log r-\log r\right]dr \\ &= \sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\log r dr+\sum_{k=1}^{n}\int_{\frac{k-1}{n}}^{\frac{k}{n}} U_n(r,\delta)\left[\log\left(\frac{k}{n}\right)-\log r\right]dr \\ &= \int_0^1 U_n(r,\delta)\log r dr+\sum_{k=1}^{n}U_{nk}\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right]dr. \end{aligned}$$

Let

$$a_k=\int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right]dr,$$

hence,

$$\frac{1}{n}\sum_{k=1}^{n} U_{nk}\log\left(\frac{k}{n}\right)=\int_0^1 U_n(r,\delta)\log r dr+\sum_{k=1}^{n}U_{nk}a_k.$$

Now,

$$\begin{aligned} a_k &= \int_{\frac{k-1}{n}}^{\frac{k}{n}}\left[\log\left(\frac{k}{n}\right)-\log r\right]dr=\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log\left(\frac{k}{n}\right)dr-\int_{\frac{k-1}{n}}^{\frac{k}{n}}\log r dr \\ &= \frac{1}{n}\log\left(\frac{k}{n}\right)-\frac{k}{n}\log\left(\frac{k}{n}\right)+\left(\frac{k-1}{n}\right)\log\left(\frac{k-1}{n}\right)+\frac{1}{n} \\ &= -\left(\frac{k-1}{n}\right)\log\left(\frac{k}{k-1}\right)+\frac{1}{n}. \end{aligned}$$

Thus, $a_1 = 1/n$. For $k > 1$, the series expansion

$$\log x=\frac{x-1}{x}+\frac{1}{2}\left(\frac{x-1}{x}\right)^2+\frac{1}{3}\left(\frac{x-1}{x}\right)^3+...$$

will be used to show that

$$\log\left(\frac{k}{k-1}\right)=\frac{1}{k}+\frac{1}{2}\left(\frac{1}{k}\right)^2+\frac{1}{3}\left(\frac{1}{k}\right)^3+...$$

and hence

$$a_k=-\left(\frac{k-1}{n}\right)\left[\frac{1}{k}+O\left(\left(\frac{1}{k}\right)^2\right)\right]+\frac{1}{n}=O\left(\frac{1}{(k-1)n}\right).$$

That is,

$$(k-1)na_k=-(k-1)^2\left[\frac{1}{k}+O\left(\left(\frac{1}{k}\right)^2\right)\right]+(k-1)=\frac{(k-1)}{k}+O\left(1\right)=O\left(1\right).$$

Given that

$$U_{nk}=O_p(1),$$

$$n\sum_{k=1}^{n}a_k\sim\sum_{k=1}^{n}\frac{1}{k-1}\sim\log n,$$

and

$$\sum_{k=1}^{n} U_{nk} a_k = O_p\left(\frac{\log n}{n}\right),$$

we have

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n} U_{nk} \log\left(\frac{k}{n}\right) &= \int_0^1 U_n(r,\delta)\log r dr + \sum_{k=1}^{n} U_{nk} a_k = \int_0^1 U_n(r,\delta)\log r dr + o_p(1) \\ &\Longrightarrow \int_0^1 \log r U_x(r,\delta) dr,\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n\log n}\sum_{k=1}^{n} U_k \log k &= \frac{1}{\log n}\left(\frac{1}{n}\sum_{k=1}^{n} U_{nk}\log\left(\frac{k}{n}\right)\right) + \frac{1}{n}\sum_{k=1}^{n} U_{nk} \\ &= \frac{1}{n}\sum_{k=1}^{n} U_{nk} + o_p(1) = \sum_{k=1}^{n}\int_{(k-1)/n}^{k/n} U_n(r,\delta) dr + o_p\left(1\right) \\ &= \int_0^1 U_n(r,\delta) dr + o_p(1) \Longrightarrow \int_0^1 U_x(r,\delta) dr.\end{aligned}$$

All together gives the stated result

$$\log n(\hat{\beta}-\beta) = \frac{\frac{1}{n\log n}\sum_{k=1}^{n} U_k \log k}{\frac{1}{n\log^2 n}\sum_{k=1}^{n}\log^2 k} \Longrightarrow \int_0^1 U_x(r,\delta) dr.$$

Q.E.D.

Proof of Proposition 5: From (1.22) and (1.23)

$$y_t - \hat{m}_t = y_t - \frac{1}{t}\sum_{j=1}^{t} y_j = x_t - \frac{1}{t}\sum_{j=1}^{t} x_j.$$

By assumption,

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{[nr]} x_t \Longrightarrow D_x(r,\delta).$$

Then, applying the CMT

$$\int_0^1 \left(\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{[nr]} x_j\right) dr \Longrightarrow \int_0^1 D_x(r,\delta) dr.$$

Therefore,

$$\begin{aligned}\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}(y_t-\hat{m}_t) &= \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_t - \frac{1}{t}\sum_{j=1}^{t} x_j\right) = \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t - \frac{1}{n}\sum_{t=1}^{n}\frac{n}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t} x_j \\ &\Longrightarrow D_x(1,\delta) - \int_0^1 r^{-1} D_x(r,\delta) dr,\end{aligned}$$

and $(y_t - \hat{m}_t) \sim S\left(\delta\right)$. **Q.E.D.**

Proof of Proposition 6: The proof will be divided in five steps.

(i) First, the partial mean is computed

$$\frac{1}{t}\sum_{j=1}^{t} y_j = m_0 + m_1 \frac{1}{t}\sum_{j=1}^{t} j + \frac{1}{t}\sum_{j=1}^{t} x_j.$$

(ii) Second, the partial mean is substracted from y_t

$$\begin{aligned} y_t - \frac{1}{t}\sum_{j=1}^{t} y_j &= m_1 t + x_t - m_1 \frac{1}{t}\sum_{j=1}^{t} j - \frac{1}{t}\sum_{j=1}^{t} x_j = m_1 t - m_1 \frac{1}{t}\frac{t\,(t+1)}{2} + x_t - \frac{1}{t}\sum_{j=1}^{t} x_j \\ &= \frac{m_1}{2}\,(t-1) + x_t - \frac{1}{t}\sum_{j=1}^{t} x_j. \end{aligned}$$

(iii) Third, compute

$$\begin{aligned} \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i\right) &= \frac{2}{t}\sum_{j=1}^{t}\left(\frac{m_1}{2}\,(j-1) + x_j - \frac{1}{j}\sum_{i=1}^{j} x_i\right) \\ &= \frac{m_1}{2}\,(t-1) + \frac{2}{t}\sum_{j=1}^{t} x_j - \frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j} x_i. \end{aligned}$$

(iv) Fourth, substracting the quantity obtained in step (iii) from that obtained in step (ii)

$$y_t - \frac{1}{t}\sum_{j=1}^{t} y_j - \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i\right) = x_t - \frac{3}{t}\sum_{j=1}^{t} x_j + \frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j} x_i.$$

(v) Finally, the asymptotic behavior of the following re-scaled sum is analyzed

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(y_t - \frac{1}{t}\sum_{j=1}^{t} y_j - \frac{2}{t}\sum_{j=1}^{t}\left(y_j - \frac{1}{j}\sum_{i=1}^{j} y_i\right)\right) = \frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n}\left(x_t - \frac{3}{t}\sum_{j=1}^{t} x_j + \frac{2}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j} x_i\right).$$

Consider the first summand. By assumption,

$$\frac{1}{n^{1/2+\delta}}\sum_{t=1}^{n} x_t \Longrightarrow D_x\,(1,\delta)\,.$$

For the second and third summands, the CMT will be used. With respect the former

$$\frac{3}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{1}{t}\sum_{j=1}^{t} x_j = \frac{3}{n}\sum_{t=1}^{n}\frac{n}{t}\frac{1}{n^{1/2+\delta}}\sum_{j=1}^{t} x_j \Longrightarrow 3\int_0^1 r^{-1} D_x\,(r,\delta)\,dr,$$

and with respect the latter

$$\frac{2}{n^{1/2+\delta}}\sum_{t=1}^{n}\frac{1}{t}\sum_{j=1}^{t}\frac{1}{j}\sum_{i=1}^{j} x_i = \frac{2}{n^2}\sum_{t=1}^{n}\frac{t^{-3/2+\delta}}{n^{-3/2+\delta}}\sum_{j=1}^{t}\frac{t}{j}\frac{1}{t^{1/2+\delta}}\sum_{i=1}^{j} x_i = o_p(1).$$

Therefore,

$$\frac{1}{n^{\frac{1}{2}+\delta}}\sum_{t=1}^{n}(y_t - \hat{m}_t) \Longrightarrow D_x\,(1,\delta) - 3\int_0^1 r^{-1} D_x\,(r,\delta)\,dr,$$

and $(y_t - \hat{m}_t) \sim S\,(\delta)$. **Q.E.D.**

Chapter 2

Co-summability

Abstract: Co-integration theory is an ideal framework to study linear relationships among persistent economic time series. Nevertheless, the intrinsic linearity in the concepts of integration and co-integration makes them unsuitable to study non-linear relationships between persistent processes. This drawback hinders the empirical analysis of modern macroeconomics which often deals with asymmetric responses to policy interventions, multiplicity of equilibria, transition between regimes or even log-linearized equilibria.

In this chapter, we formalize the idea of *co-summability*, which is built upon the concept *order of summability* introduced in the previous chapter and conceived to deal with non-linear transformations of persistent processes. Theoretically, a co-summable relationship is balanced and describes a long run equilibrium that can be non-linear. To test for these type of equilibria, inference tools for balancedness and co-summability are designed and their asymptotic properties are analyzed. The finite sample performance is studied through Monte Carlo experiments.

The practical strength of co-summability theory is shown through two empirical applications. Specifically, the hypothesis of asymmetric preferences of central bankers and the environmental Kuznets curve are studied through the lens of co-summability.

2.1 Introduction

Co-integration theory has received a great deal of attention from economists and econometricians. From a theoretical point of view, co-integration played the role of combining properly persistent economic time series with linear long run equilibrium relationships hypothesized by economic theorists. In the economic literature, co-integration meant a positive step towards consensus in the historically "*measurement without theory*" vs "*theory without measurement*" debate. Economic theories implying co-integrating relationships among economic time series contributed to give this step. From an empiricist point of view, co-integration came up with a clear and precise applied methodology to estimate and test for those economic hypothesis.

To provide richer descriptions of the economic phenomena, researchers have gotten into the non-linear world. Nevertheless, the ideas of integration and co-integration cannot be directly used to analyze non-linear equilibrium relationships among persistent variables since these concepts do not properly apply. To be more precise, consider the following non-linear relationship: $y_t = f(x_t, \theta) + u_t$. If it were known that $f(x_t, \theta)$ is $I(d)$, then the standard framework of co-integration could fit perfectly. However, when x_t is persistent, say $I(1)$, then for many interesting non-linear transformations f the order of integration of $f(x_t, \theta)$ may not be well defined. This failure of applicability of the definition of order of integration has two important drawbacks. First, it is not possible to know whether the postulated relationship is balanced –a necessary, although not sufficient, condition for having correctly specified the model. Second, the concept of co-integration cannot be directly extended to non-linear long run relationships. These two consequences originate a clear need for theoretically valid and empirically useful concepts that generalize those of integration and co-integration.

This chapter proposes to use the idea of order of summability formalized in Chapter 1. It was conceived to deal both theoretically and empirically with non-linear transformations of persistent processes. Making use of this new concept, co-integration theory can be generalized by defining balancedness and co-summability. In addition, by taking advantage of the order of summability estimator, balancedness and co-summability can be empirically analyzed.

To infer if a postulated relationship is balanced, the rate of convergence estimator in McElroy and Politis (2007) and subsampling techniques can be used. Once balancedness is achieved, researchers must distinguish between spurious or co-summable regressions. This chapter proposes a residual based test to disentangle that question; therefore, an estimate of the errors is needed. Parametric and non-parametric approaches to estimate non-linear long run relationships are available in the literature. Park and Phillips (2001) and Wang and Phillips (2009) develop parametric and non-parametric methods, respectively, from an integrated processes perspective. Alternatively, Karlsen, Myklebust and Tjϕstheim (2007) and Schienle (2011) analyze nonparametric estimation in a recurrent Markov chains setup. Notwithstanding, all these studies assume that the regression model specifies a co-integrating relation; something that should be tested in practice. There has been some, rather limited, proposals in this direction –see, for instance, Choi and Saikkonen (2010). In this chapter, parametric regression models that are non-linear in variables but linear in parameters will be taken into consideration. In this scenario, the asymptotic properties of the ordinary least squares estimator under unbalancedness, spuriousness and co-summability are studied. These properties guarantee being able to discriminate between spurious or co-summable regressions through a residual based test.

Finally, these tools are put into practice with two different empirical applications where non-linear transformations of persistent processes arise. Specifically, asymmetric preferences of central bankers and the environmental Kuznets curve are studied using co-summability theory. The former hypothesis is translated, in the literature, into non-linear Taylor rules when conducting monetary policy –see, for instance, Clarida and Gertler (1997) or Dolado, María-Dolores and Naveira (2005). These non-linearities and the fact that the variables involved in this type of rules are found to be persistent, makes co-summability appropriate to be used in this context. The latter hypothesis, the environmental Kuznets curve, postulates an inverted U-shaped relationship between pollution and economic development, usually measured by CO_2 emissions and GDP, respectively. Again, this non-linear relationship jointly with the well documented persistence of these two measures makes this hypothesis another natural economic context where co-summability theory rightly fits. The corresponding empirical results give new insights for the econometric treatment of these two hypothesis. In the Taylor rule case, the linear specification does not define a long run relationship, pointing to a possible misspecification. Following the asymmetric preference of central bankers literature, we find that a threshold Taylor rule describes an equilibrium in the long run. Specifically, it is found that the Federal Reserve reacts asymmetrically to recessions and expansions in an aggressive way. With respect the environmental Kuznets curve, favorable evidence is found when the logarithms of the time series involved are analyzed. However, when the levels of these series are studied an unbalanced relationship is found. Therefore, as a practical recommendation to this literature, logarithms of CO_2 emissions and GDP should be considered when studying polynomial reduced forms of the environmental Kuznets curve in order to get consistent estimates of the parameters of the model.

The chapter is organized as follows. In Section 2.2, balancedness and co-summability are formally defined and discussed through some economic examples. Section 2.3 develops an empirical strategy to test for co-summability. First, a test for balancedness is designed. Then, a test for co-summability is proposed. The finite sample performance of these procedures is studied via simulations. In Section 2.4, the proposed tools are applied to test for two economic hypothesis: asymmetric preferences of central bankers and the environmental Kuznets curve. Finally, Section 2.5 finishes with some concluding remarks. All the proofs are collected in the Appendix at the end of the chapter.

A word on notation. We use the symbol "$\Longrightarrow$" to signify convergence in distribution and weak convergence indistinctly, "$\xrightarrow{p}$" to signify convergence in probability. Stochastic processes such as the standard Brownian motion $W(r)$ are defined on $[0,1]$. Finally, all limits given in this chapter are taken as the sample size $n \to \infty$.

2.2 Balancedness and Co-summability

2.2.1 Order of Summability

The subsequent theory relies on the idea of order of summability of stochastic processes. It was first introduced in a heuristic way in Gonzalo and Pitarakis (2006) and subsequently formalized in Chapter 1 of this thesis.

Definition 5 *: A stochastic process y_t with positive variance is said to be summable of order δ, represented as $S(\delta)$, if*

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1) \qquad \text{as } n \to \infty,$$

where δ is the minimum real number that makes S_n bounded in probability, m_t is a deterministic sequence, and $L(n)$ is a slowly-varying function[8]*.*

The order of summability, δ, gives a summary measure of the stochastic properties –persistence and evolution of the variance– of y_t without relying on a particular data generating process. In this sense, it can overcome the drawbacks of using the order of integration in non-linear environments.

Let

$$\pi_t = \pi_{t-1} + \varepsilon_t, \tag{2.1}$$

with $\pi_0 = 0$ and $\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2)$.

Example 8 *: Square of a random walk*

Consider

$$\pi_t^2 = \pi_{t-1}^2 + 2\pi_{t-1}\varepsilon_t + \varepsilon_t^2. \tag{2.2}$$

[8] A positive, Lebesgue measurable function L, on $(0,\infty)$ is slowly varying –in the Karatama's sense– at ∞ if

$$\frac{L(\lambda n)}{L(n)} \to 1 \;\; (n \to \infty) \; \forall \lambda > 0.$$

(See Embrechts, Klüppelberg and Mikosh, 1999, p.564).

To establish the order of integration of this process is not an obvious task. Granger (1995) considers that π_t^2 can be seen as a random walk with drift, hence, one could think that π_t^2 is $I(1)$. However,

$$V[\pi_t^2 - \pi_{t-1}^2] = 4(t-1)\sigma_\varepsilon^4 + E[\varepsilon_t^4].$$

In fact, this process can be though as having an infinite order of integration since the variance of $\Delta^d \pi_t^2$ depends on t regardless of the values of d. Nevertheless, it is well known that

$$S_n = \frac{1}{n^2\sigma_\varepsilon^2}\sum_{t=1}^{n}\pi_t^2 \Longrightarrow \int_0^1 W^2(r)dr.$$

Hence, π_t^2 is $S(1.5)$.

Example 9 *: Product of Indicator Function and Random Walk*

Let

$$h_t = 1(v_t \leq \gamma)\pi_t, \tag{2.3}$$

where $v_t \sim i.i.d.\,(0,1)$ is independent of ε_t and $1(\cdot)$ is the indicator function. This is another example where the concept of integrability is difficult to apply. Strictly speaking, h_t has an infinite order of integration but this is not a useful characterization in practice. Instead, the order of summability can be easily obtained. Given that

$$S_n = \frac{1}{n^{\frac{3}{2}} p\sigma_\varepsilon}\sum_{t=1}^{n} h_t \Longrightarrow \int_0^1 W(r)dr,$$

where $p = \Pr(v_t \leq \gamma)$, h_t is $S(1)$.

From a multivariate perspective, an applied economist starts its analysis from a postulated economic relationship, say $y_t = f(x_t, \theta)$. Then, recognizing that it is just an approximation to reality and θ is typically unknown, the difference $u_t = y_t - f(x_t,\theta)$ is statistically analyzed.

Assumption 0.

$$S_{yn} = \frac{1}{n^{1/2+\delta_y}}\sum_{t=1}^{n} y_t \Longrightarrow D_y \quad \text{and} \quad S_{zn} = \frac{1}{n^{1/2+\delta_z}}\sum_{t=1}^{n} f(x_t,\theta) \Longrightarrow D_z,$$

where D_y and D_z are two random variables with positive variance.

Under Assumption 0, $y_t \sim S(\delta_y)$ and $f(x_t,\theta) \sim S(\delta_z)$. This assumption will be particularly convenient to put forward the balancedness of a theoretical hypothesis.

2.2.2 Balancedness

Definition 6 *: A postulated relationship*

$$y_t = f(x_t, \theta),$$

will be said to be balanced if $y_t \sim S(\delta_y)$, $z_t = f(x_t,\theta) \sim S(\delta_z)$, and $\delta_y = \delta_z$.

Given a theoretical hypothesis

$$y_t = f(x_t, \theta), \tag{2.4}$$

the order of summability of x_t, δ_x, could differ from that of $z_t = f(x_t, \theta)$, δ_z. This means that given δ_y and δ_x there will be only some appropriate transformations $f(., \theta)$ that will generate balanced relationships, i.e. $\delta_y = \delta_z$. This is not only important for econometricians but also for economic theorists when choosing functional forms to construct their theories.

Notice that under Assumption 0, an unbalanced postulated model is clearly misspecified –in a wide sense. When $\delta_y > \delta_z$,

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} f(x_t, \theta) = o_p(1),$$

which is not true when Assumption 0 holds. On the other hand, if $\delta_y < \delta_z$,

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} f(x_t, \theta),$$

with the right hand side being unbounded. Again, an incongruity with Assumption 0. Hence, balancedness becomes a necessary, although not sufficient, condition for a correct specification. Particular economic examples will show the relevance of balancedness in practice.

Example 10 *: Endogenous Growth Models (Jones, 1995)*

Endogenous growth theory implies that permanent changes in policy variables such as the investment rate in physical capital have permanent effects on the rate of economic growth. The equation of interest is

$$g_{yt} = -\ddot{\delta} + \tilde{A} i_{kt}, \tag{2.5}$$

where g_{yt} is the growth rate of the economy, $\ddot{\delta}$ is the rate of depreciation, $\tilde{A}$ measures the total factor productivity, and i_{kt} is the investment rate in physical capital. If this equation was balanced, then the persistence of the growth rate should be similar to that of the investment rate. Nevertheless, using time series techniques it is found that U.S. growth rates exhibit no large persistent changes while large and permanent movements are found in investment rates. Hence, Jones (1995) argues that endogenous growth models are rejected by this criterion. The balancedness of a postulated equation was already an important feature in the linear modelling.

Balancedness will be particularly important in non-linear models involving persistent variables. As already stated in Granger (1995), non-linear transformations of heterogeneous and persistent processes can have an important impact on their stochastic properties. This impact could be hardly contemplated by the order of integration but can be asserted by the order of summability. The next two examples illustrate the importance of balancedness in non-linear economic models.

Example 11 : *Central Bankers with Asymmetric Preferences*

Consider a central bank with asymmetric preferences respect to deviations of inflation or output from some particular target level. Under such preferences the central bank would react more or less aggressively when inflation or output deviates from above than from below the target. Different modelizations of this hypothesis based on Taylor rules can be found in the literature. For instance, Clarida and Gertler (1997) study the following threshold type Taylor rule for the Bundesbank

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t 1\left(\tilde{\pi}_t > 0\right) + \theta_2 \tilde{\pi}_t 1\left(\tilde{\pi}_t \leq 0\right) + \theta_3 \tilde{y}_t 1\left(\tilde{\pi}_t > 0\right) + \theta_4 \tilde{y}_t 1\left(\tilde{\pi}_t \leq 0\right), \tag{2.6}$$

where i_t denotes interest rates, $\tilde{\pi}_t$ are deviations from inflation target, and $\tilde{y}_t$ is the output gap. On the other hand, Dolado, María-Dolores and Naveira (2005) allowing for a nonlinear Phillips curve derive the following optimal monetary policy rule

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t + \theta_2 \tilde{y}_t + \theta_3 \tilde{\pi}_t \tilde{y}_t. \tag{2.7}$$

In both cases, studying balancedness of these equations will be troublesome using the $I(d)$ framework. Even if it can be said that i_t, $\tilde{\pi}_t$, and $\tilde{y}_t$ are $I(d_i)$, $I(d_{\tilde{\pi}})$, and $I(d_{\tilde{y}})$, respectively, the order of integration of $\tilde{\pi}_t 1\left(\tilde{\pi}_t \geq 0\right)$ or $\tilde{\pi}_t \tilde{y}_t$ could not be well defined. Nevertheless, the generality of the order of summability makes it suitable to be used in both situations.

Example 12 : *Environmental Kuznets Curve*

The environmental Kuznets curve points an inverted-U relationship between pollution and economic development –see Dasgupta et al. (2001) or Brock and Taylor (2005) for an overview. The usual shape given to this relationship is of a polynomial type. Consider the simplest

$$p_t = \theta_0 + \theta_1 y_t + \theta_2 y_t^2,$$

where p_t is a measure of pollution and y_t is a measure of income, typically CO_2 and GDP, respectively. Again, to check whether this equation is balanced will be troublesome if the order of integration is used. Even if it is known that y_t is $I\left(d_y\right)$, the order of integration of y_t^2 could not be well defined. As it has been emphasized above, the order of summability can help to overcome this pitfall. In fact, the polynomial order in the environmental Kuznets curve is a controversy in the literature. The order of summability is an objective criterion to determine it. Since it does not rely on any particular structure of the data generating process, it is suitable to be generally used.

2.2.3 Co-summability

Definition 7 : *Two summable stochastic processes, $y_t \sim S\left(\delta_y\right)$ and $x_t \sim S\left(\delta_x\right)$, will be said to be co-summable if there exists $z_t = f\left(x_t, \theta\right) \sim S\left(\delta_y\right)$ such that $u_t = y_t - f(x_t, \theta)$ is $S(\delta_u)$, with $\delta_u = \delta_y - \delta$ and $\delta > 0$. In short, $(y_t, z_t) \sim CS(\delta_y, \delta)$.*

Some aspects of this definition are worth to mention. Firstly, even when x_t is $S(\delta_x)$ with $\delta_x > 0$, some functions f can make $f(x_t, \theta) \sim S(0)$. As in co-integration theory, relations in which y_t and $f(x_t, \theta)$ are $S(0)$ will be excluded from our co-summability analysis. Notwithstanding, it should be emphasized the relevance of this type of relationships. They allow to relate persistent and non-persistent time series, such as growth rates and levels of macroeconomic time series in the long run; although in a non-linear fashion. These relations deserve further research outside the co-summability framework.

Secondly, a co-summable relationship is balanced. As already stated, balancedness is a necessary, although not sufficient, condition for a correct specification. In fact, when $\delta_y = \delta_z$, the relationship $y_t = f(x_t, \theta)$ could be balanced spuriously. As in standard co-integration theory, spuriousness and co-summability can be distinguished through the fact that only under co-summability $\delta_u < \delta_y$, highlighting the existence of an attractor to the equilibrium relationship.

Finally, it is important to emphasize that co-summability mimics the idea of co-integration. This fact facilitates the development of an empirical strategy to test for co-summability that inherits the steps of testing for co-integration, although uses new econometric tools.

Table 6 summarizes the four possible situations that can arise from the different configurations of orders of summability.

Table 6

Unbalancedness		Balancedness	
U1	$\delta_y > \delta_z$	**S**	$\delta_y = \delta_z$
U2	$\delta_y < \delta_z$	**C**	$\delta_y = \delta_z$, $\delta_u < \delta_y$

$U1$: Unbalancedness of type 1; $U2$: Unbalancedness of type 2; S: Spuriousness; and C: Co-summability.

2.3 Estimation and Inference

2.3.1 The model

Consider the following least squares regression

$$y_t = \hat{\theta} f(x_t) + \hat{u}_t, \tag{2.8}$$

where $f : \mathbb{R} \to \mathbb{R}$, x_t, and y_t are known by the researcher. For the shake of simplicity, only the bivariate case (y_t, x_t) will be considered. The extension to a multivariate x_t can be easily adapted.

The key aspects of our analysis are the single equation framework and the linearity in parameters of the model.

Co-summable processes share an equilibrium relationship in the long run, i.e. an attractor that can be linear or not. Applied researchers will be interested in estimating and testing for those types of relationships. Following exactly the same logic of co-integration theory, the following empirical strategy is devised.

Step 1. Test $H_o : \delta_y = \delta_z$. If it is not rejected, then go to Step 2.

Step 2. Test $H_o : \delta_{\hat{u}} = 0$.

2.3.2 Testing for Balancedness

To establish balancedness in practice we propose to start estimating the orders of summability of y_t and $z_t = f(x_t)$. To accomplish that requirement, the order of summability estimator developed in Chapter 1 can be used. It follows the convergence rate estimation procedure in McElroy and Politis (2007), which is based on a simple least squares regression. The procedure requires the following assumption.

Assumption 1. $P(S_n = 0) = 0$ *for all* $n = 1, 2, 3, ...$

Our proposal to estimate δ, which follows the convergence rate estimator in McElroy and Politis (2007), is based on the transformation $U_n = \log S_n^2$. Hence, Assumption 1 is needed to avoid taking the logarithm of zero. Specifically, under Assumptions 0 and 1,

$$U_{yk} = \log S_{yk}^2 = \log \left[\left(\frac{1}{k^{\frac{1}{2}+\delta_y}} \sum_{t=1}^{k} y_t \right)^2 \right] = O_p(1),$$

and the following regression model can be derived

$$Y_{yk} = \beta_y \log k + U_{yk}, \tag{2.9}$$

where $Y_{yk} = \log \left(\sum_{t=1}^{k} y_t \right)^2$ and $\beta_y = 1 + 2\delta_y$.

Chapter 1 shows that the OLS estimator of $\beta_y = 1 + 2\delta_y$ is $\log n$-consistent with an asymptotic distribution that cannot be tabulated in general. Through simulations, it is shown that subsampling confidence intervals can be constructed to undertake inferences on the true δ_y. It is important to mention that the presence of deterministic components in the DGP has a strong bias effect on the order of summability estimator, at least in finite samples. In Chapter 1 valid demeaning and detrending procedures are developed. Nevertheless, to facilitate exposition no deterministic components will be considered in this section.

Notice that the regression to estimate the order of summability of z_t is

$$Y_{zk} = \beta_z \log k + U_{zk}, \tag{2.10}$$

where $Y_{zk} = \log \left(\sum_{t=1}^{k} z_t \right)^2$ and $\beta_z = 1 + 2\delta_z$.

To test for balancedness, an auxiliary equation that substracts (2.10) from (2.9) will be used, that is,

$$Y_{yk} - Y_{zk} = \left(\beta_y - \beta_z \right) \log k + U_{yk} - U_{zk}.$$

Let $Y_k = Y_{yk} - Y_{zk}$, $\beta = \beta_y - \beta_z$, and $U_k = U_{yk} - U_{zk}$. Then, testing $H_o : \delta_y = \delta_z$ is equivalent to test $H_o : \beta = 0$ in

$$Y_k = \beta \log k + U_k. \tag{2.11}$$

Proposition 7 *: Let $\hat{\beta}_n$ be the ordinary least squares estimator of β in (2.11). If*

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{[nr]} y_t \Longrightarrow D_y \left(r, \delta_y \right) \quad \text{and} \quad \frac{1}{n^{1/2+\delta_z}} \sum_{t=1}^{[nr]} z_t \Longrightarrow D_z \left(r, \delta_z \right),$$

where $D_y \left(r, \delta_y \right)$ and $D_z \left(r, \delta_z \right)$ are two random processes with positive variance, then

$$\log n \left(\hat{\beta}_n - \beta \right) \Longrightarrow \int_0^1 \left(U_y \left(r, \delta_y \right) - U_z \left(r, \delta_z \right) \right) dr,$$

where $U_y \left(r, \delta_y \right) = \log \left[\left(r^{-1/2-\delta_y} D_y \left(r, \delta_y \right) \right)^2 \right]$ and $U_z \left(r, \delta_z \right) = \log \left[\left(r^{-1/2-\delta_z} D_z \left(r, \delta_z \right) \right)^2 \right]$.

Remark: Proposition 7 shows that $\hat{\beta}_n$ is a consistent estimator of the difference $\beta_y - \beta_z$. In particular, under balancedness

$$\hat{\beta}_n \xrightarrow{p} 0.$$

Nevertheless, the asymptotic distribution cannot be tabulated in general. As in Chapter 1, we use subsampling confidence intervals to undertake inference. Their finite sample performance is analyzed via Monte Carlo experiments.

In line with Table 6, the four situations of interest are *S: spuriousness, C: co-summability, U1: unbalancedness of type 1, and U2: unbalancedness of type 2.* Let $x_{yt} = x_{y,t-1} + \varepsilon_{yt}$ with $\varepsilon_{yt} \sim i.i.d.N\left(0,1\right)$ and $x_{y0} = 0$. $x_{zt} = x_{z,t-1} + \varepsilon_{zt}$ with $\varepsilon_{zt} \sim i.i.d.N\left(0,1\right)$ and $x_{z0} = 0$. In addition, let $u_t \sim i.i.d.N\left(0,1\right)$ and $v_t \sim i.i.d.N\left(0,1\right)$. ε_{yt}, ε_{zt}, u_t, and v_t are independent of each other. We study the data generating processes –DGP– collected in Table 7.

In all cases $\hat{\beta}_n$ is calculated. Then, the subsampling confidence interval is computed and the null hypothesis of balancedness, $H_o : \beta = 0 \equiv \delta_y - \delta_z = 0$, is analyzed. Performance is measured by coverage probability of two-sided nominal 95% symmetric intervals for the null hypothesis. The experiment is based on 1000 replicas and three different sample sizes $n = \{100,\ 500,\ 1000\}$. A subsample size $b = \sqrt{n}$ has been chosen. Results are collected in Table 8.

Table 7: DGPs for Monte Carlo Experiments

S	y_t	z_t	**C**	y_t	z_t
1	$\ln(\lvert x_{yt}\rvert)$	$\ln(\lvert x_{zt}\rvert)$	1	z_t+u_t	$\ln(\lvert x_{zt}\rvert)$
2	$v_t x_{yt}$	$v_t x_{zt}$	2	z_t+u_t	$v_t x_{zt}$
3	$\Delta^{0.25} x_{yt}$	$\Delta^{0.25} x_{zt}$	3	z_t+u_t	$\Delta^{0.25} x_{zt}$
4	x_{yt}	x_{zt}	4	z_t+u_t	x_{zt}
5	$1(v_t \leq 0)\, x_{yt}$	$1(v_t \leq 0)\, x_{zt}$	5	z_t+u_t	$1(v_t \leq 0)\, x_{zt}$
6	x_{yt}^2	x_{zt}^2	6	z_t+u_t	x_{zt}^2
7	$\sum_{j=1}^{t} x_{yj}$	$\sum_{j=1}^{t} x_{zj}$	7	z_t+u_t	$\sum_{j=1}^{t} x_{zj}$
U1	y_t	z_t	**U2**	y_t	z_t
1	$v_t x_{yt}$	ε_{zt}	1	ε_{yt}	$v_t x_{zt}$
2	x_{yt}	ε_{zt}	2	ε_{yt}	x_{zt}
3	x_{yt}^2	ε_{zt}	3	ε_{yt}	x_{zt}^2
4	$\sum_{j=1}^{t} x_{yj}$	ε_{zt}	4	ε_{yt}	$\sum_{j=1}^{t} x_{zj}$

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1 and unbalancedness of type 2, respectively. $x_{yt} = x_{y,t-1} + \varepsilon_{yt}$ with $\varepsilon_{yt} \sim i.i.d.N(0,1)$ and $x_{y0} = 0$. $x_{zt} = x_{z,t-1} + \varepsilon_{zt}$ with $\varepsilon_{zt} \sim i.i.d.N(0,1)$ and $x_{z0} = 0$. In addition, let $u_t \sim i.i.d.N(0,1)$ and $v_t \sim i.i.d.N(0,1)$. ε_{yt}, ε_{zt}, u_t, and v_t are independent of each other.

It is worth to mention that these experiments study through coverage probabilities the performance of the test in terms of size and power. Hence, under the null hypothesis of balancedness, a coverage probability of 95% should be expected as a measure of the 95% nominal level of significance. Under the alternative hypothesis, a low coverage probability implies a high power of the test.

As it can be seen in Table 8, under the null hypothesis –S and C– a high coverage probability, higher than the 95% nominal confidence level, is often obtained. This means that the test is slightly undersized, leading to an over non-rejection of the null hypothesis. The implication is a high probability to jump to Step 2 –testing for co-summability– in the proposed empirical strategy. Under the alternative hypothesis –U1 and U2– low coverage probabilities should be expected. In these cases, for a given sample size, results show that the higher the difference $\delta_y - \delta_z$ in absolute value, the lower the coverage probability. Furthermore, under the alternative hypothesis, for a given DGP, the greater the sample size the lower the coverage probability. In other words, as expected,

Table 8: Coverage Probabilities. Testing for Balancedness: $Ho : \delta_y = \delta_z$

S			n			**C**			n		
DGP	δ_y	δ_z	100	500	1000	DGP	δ_y	δ_z	100	500	1000
1	1/2	1/2	0.940	1.000	1.000	1	1/2	1/2	1.000	1.000	1.000
2	1/2	1/2	0.983	0.993	0.995	2	1/2	1/2	0.998	1.000	0.999
3	3/4	3/4	0.937	0.974	0.978	3	3/4	3/4	1.000	1.000	1.000
4	1	1	0.944	0.972	0.967	4	1	1	1.000	1.000	1.000
5	1	1	0.798	0.839	0.825	5	1	1	0.998	1.000	1.000
6	3/2	3/2	0.996	0.999	1.000	6	3/2	3/2	0.998	0.990	0.993
7	2	2	0.954	0.963	0.975	7	2	2	0.993	0.948	0.919
U1			n			**U2**			n		
DGP	δ_y	δ_z	100	500	1000	DGP	δ_y	δ_z	100	500	1000
1	1/2	0	0.891	0.821	0.791	1	0	1/2	0.901	0.818	0.802
2	1	0	0.362	0.163	0.098	2	0	1	0.400	0.160	0.094
3	3/2	0	0.177	0.041	0.011	3	0	3/2	0.202	0.031	0.013
4	2	0	0.169	0.032	0.004	4	0	2	0.205	0.024	0.002

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1, and unbalancedness of type 2, respectively.

the power of the test improves as we move far away from the null hypothesis and the sample size increases. It is important to mention that these results are sensible to the co-summability vector. A modified version of the above procedure to eliminate this dependence is being developed.

2.3.3 Asymptotic Properties of $\hat{\theta}$

In this section, the asymptotic properties of the OLS estimator

$$\hat{\theta} = \frac{\sum_{t=1}^{n} y_t f(x_t)}{\sum_{t=1}^{n} f^2(x_t)}.$$

under the four situations of interest –U1, U2, S, and C– are studied.

2.3.3.1 Unbalanced and Spurious Relationships

As in cointegration theory, regressions involving unbalanced and spurious relationships will be understood as those in which y_t and z_t are independent of each other. Their distinctive characteristic is the relationship between its orders of summability.

To study the asymptotic properties of $\hat{\theta}$ under unbalancedness and spuriousness the following assumption will be made.

Assumption 2. Let $(y_{nt}, z_{nt}) = (y_t/n^{\alpha_y}, z_t/n^{\alpha_z}) = (y_t/n^{\alpha_y}, f(x_t)/n^{\alpha_z})$*. The D-space analog of* (y_{nt}, z_{nt}) *satisfies*

$$(y_n(r, \alpha_y), z_n(r, \alpha_z)) = \left(\frac{y_{[nr]}}{n^{\alpha_y}}, \frac{f\left(x_{[nr]}\right)}{n^{\alpha_z}}\right) \Longrightarrow (D_y(r, \alpha_y), D_z(r, \alpha_z)).$$

The relationship between Assumption 2 and the order of summability of y_t and z_t follows directly from applying the Continuous Mapping Theorem –CMT. That is,

$$\left(\frac{1}{n}\sum_{t=1}^{n}\frac{y_t}{n^{\alpha_y}}, \frac{1}{n}\sum_{t=1}^{n}\frac{z_t}{n^{\alpha_z}}\right) = \left(\int_0^1 \frac{y_{[nr]}}{n^{\alpha_z}}dr, \int_0^1 \frac{z_{[nr]}}{n^{\alpha_z}}dr\right) \Longrightarrow \left(\int_0^1 D_y(r, \alpha_y)\,dr, \int_0^1 D_y(r, \alpha_y)\,dr\right),$$

which implies that $\delta_y = 1/2 + \alpha_y$ and $\delta_z = 1/2 + \alpha_z$.

Proposition 8 *: Under Assumption 2, if* y_t *and* z_t *are independent,*

$$n^{\delta_z - \delta_y}\hat{\theta} \Longrightarrow \frac{\int_0^1 D_z(r, \alpha_z)\, D_y(r, \alpha_y)\,dr}{\int_0^1 D_z^2(r, \alpha_z)\,dr}.$$

Remark: Under unbalancedness of type 1, $\delta_z - \delta_y < 0$. Hence, $\hat{\theta}$ diverges. In a spurious relationship, $\delta_z - \delta_y = 0$. Therefore, $\hat{\theta}$ converges in distribution to a random variable. Finally, under unbalancedness of type 2, $\delta_z - \delta_y > 0$. Hence, $\hat{\theta}$ converges in probability to zero in this case.

2.3.3.2 Co-summable Relationships

Let

$$y_t = \theta f(x_t) + u_t.$$

Assumption 3. Weak Co-summability. Let $(z_{nt}, u_{nt}) = (z_t/n^{\alpha_z}, u_t/n^{\alpha_u}) = (f(x_t)/n^{\alpha_z}, u_t/n^{\alpha_u})$*. The D-space analog of* (z_{nt}, u_{nt}) *satisfies*

$$(z_n(r, \alpha_z), u_n(r, \alpha_u)) = \left(\frac{f\left(x_{[nr]}\right)}{n^{\alpha_z}}, \frac{u_{[nr]}}{n^{\alpha_u}}\right) \Longrightarrow (D_z(r, \alpha_z), D_u(r, \alpha_u)),$$

where $\alpha_u < \alpha_y = \alpha_z$.

By the CMT, under weak co-summability $u_t \sim S(\delta_u)$ with $\delta_u = 1/2 + \alpha_u$. Given that $\alpha_u < \alpha_y = \alpha_z$, $\delta_u < \delta_y = \delta_z$ as required by Definition 7.

Proposition 9 *: Under Assumption 3, if* z_t *and* u_t *are independent*

$$n^{\delta_z - \delta_u}\left(\hat{\theta} - \theta\right) \Longrightarrow \frac{\int_0^1 D_z(r, \alpha_z)\, D_u(r, \alpha_u)\,dr}{\int_0^1 D_z^2(r, \alpha_z)\,dr}.$$

Now, let

$$v_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t,$$

and set $\mathcal{F}_t = \sigma(z_{ni}, u_i : i \leq t, n \geq 1)$ to be the smallest sigma-field containing the past history of (z_{nt}, u_t) for all n, and denote $E(X|\mathcal{F}_i)$ by $E_i X$.

Assumption 4. Strong Co-summability. $(z_n(r, \alpha_z), v_n(r)) \Longrightarrow (D_z(r, \alpha_z), D_u(r))$.

Proposition 10 *: Under Assumption 4 if*

$$(a) \sup_{i \leq n} |u_i| < \infty,$$

and

$$(b) \sup_{i \leq n+1} \sum_{k=1}^{\infty} |E_{i-1} u_{i-1+k}| < \infty,$$

then

$$n^{\delta_z}\left(\hat{\theta} - \theta\right) \Longrightarrow \frac{\int_0^1 D_z(r, \alpha_z)\, dD_u(r)}{\int_0^1 D_z^2(r, \alpha_z)\, dr}.$$

Remark: Since co-summability mimics co-integration theory, it is not surprising that for a co-summable relationship, $y_t = \theta f(x_t) + u_t$, the OLS estimator $\hat{\theta}$ is consistent and its rate of convergence depends on the difference $\delta_z - \delta_u$.

2.3.3.3 Examples

(a) Unbalanced Relationship of Type 1, $\boldsymbol{\delta}_y > \boldsymbol{\delta}_z$: Let $y_t = y_{t-1} + \varepsilon_{yt}$ and $x_t = x_{t-1} + \varepsilon_{yt}$, with $y_0 = x_0 = 0$ and $\varepsilon_{yt} \sim i.i.d.(0,1)$ independent of $\varepsilon_{xt} \sim i.i.d.(0,1)$. Consider now the following OLS regression

$$y_t = \hat{\theta} |x_t|^{1/2} + \hat{u}_t.$$

In this case, $y_t \sim S(1)$ and $z_t = |x_t|^{1/2} \sim S(3/4)$. The OLS estimator $\hat{\theta}$ satisfies

$$\frac{1}{n^{1/4}} \hat{\theta} = \frac{\frac{1}{n^{7/4}} \sum_{t=1}^{n} y_t |x_t|^{1/2}}{\frac{1}{n^{3/2}} \sum_{t=1}^{n} x_t} \Longrightarrow \frac{\int_0^1 W_y(r) |W_x(r)|^{1/2} dr}{\int_0^1 |W_x(r)|\, dr}.$$

Hence, $\hat{\theta}$ diverges as $n \to \infty$.

(b) Unbalanced Relationship of Type 2, $\boldsymbol{\delta}_y < \boldsymbol{\delta}_z$: Let $y_t = y_{t-1} + \varepsilon_{yt}$ and $x_t = x_{t-1} + \varepsilon_{yt}$, with $y_0 = x_0 = 0$ and $\varepsilon_{yt} \sim i.i.d.(0,1)$ independent of $\varepsilon_{xt} \sim i.i.d.(0,1)$. Consider the OLS regression

$$y_t = \hat{\theta} x_t^2 + \hat{u}_t.$$

In this case, $y_t \sim S(1)$ and $z_t = x_t^2 \sim S(3/2)$, and

$$\sqrt{n}\hat{\theta} = \frac{\frac{1}{n^{5/2}} \sum_{t=1}^{n} y_t x_t^2}{\frac{1}{n^3} \sum_{t=1}^{n} x_t^4} \Longrightarrow \frac{\int_0^1 W_y(r) W_x^2(r) dr}{\int_0^1 W_x^4(r) dr}.$$

Therefore, $\hat{\theta} \xrightarrow{p} 0$.

(c) Spurious Relationship, $\boldsymbol{\delta}_y = \boldsymbol{\delta}_z = \boldsymbol{\delta}_u$: Let $y_t = y_{t-1} + \varepsilon_{yt}$ and $x_t = x_{t-1} + \varepsilon_{yt}$, with $y_0 = x_0 = 0$ and $\varepsilon_{yt} \sim i.i.d.(0,1)$ independent of $\varepsilon_{xt} \sim i.i.d.(0,1)$. Consider the following OLS regression

$$y_t = \hat{\theta}_1 1(v_t \leq \gamma) x_t + \hat{\theta}_2 1(v_t > \gamma) x_t + \hat{u}_t.$$

where $v_t \sim i.i.d.(0,1)$ is independent of ε_{yt} and ε_{xt}. In this case, $y_t \sim S(1)$, $z_{1t} = 1(v_t \leq \gamma) x_t \sim S(1)$ and $z_{2t} = 1(v_t > \gamma) x_t \sim S(1)$. The OLS estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ satisfy

$$\hat{\theta}_1 = \frac{\frac{1}{n^2} \sum_{t=1}^{n} y_t 1(v_t \leq \gamma) x_t}{\frac{1}{n^2} \sum_{t=1}^{n} 1(v_t \leq \gamma) x_t^2} \Longrightarrow \frac{\int_0^1 W_y(r) W_x(r) dr}{\int_0^1 W_x^2(r) dr},$$

and

$$\hat{\theta}_2 = \frac{\frac{1}{n^2} \sum_{t=1}^{n} y_t 1(v_t > \gamma) x_t}{\frac{1}{n^2} \sum_{t=1}^{n} 1(v_t > \gamma) x_t^2} \Longrightarrow \frac{\int_0^1 W_y(r) W_x(r) dr}{\int_0^1 W_x^2(r) dr}.$$

Therefore, $\hat{\theta}_1$ and $\hat{\theta}_2$ tend to a random variable as $n \to \infty$.

(d) Co-summable Relationship, $\boldsymbol{\delta}_y = \boldsymbol{\delta}_z > \boldsymbol{\delta}_u$: Let $x_t = x_{t-1} + \varepsilon_{yt}$, with $x_0 = 0$ and $\varepsilon_{xt} \sim i.i.d.(0,1)$. Consider now the following long run relationship,

$$y_t = \theta_1 1(v_t \leq \gamma) x_t + \theta_2 1(v_t > \gamma) x_t + u_t,$$

where $u_t \sim i.i.d.(0,1)$ is independent of ε_{xt}, and u_t and ε_{xt} are independent of $v_t \sim i.i.d.(0,1)$. In this case, $z_{1t} = 1(v_t \leq \gamma) x_t \sim S(1)$, $z_{2t} = 1(v_t > \gamma) x_t \sim S(1)$ and $u_t \sim S(0)$. Therefore, $y_t \sim S(1)$. The OLS estimator of θ_1 and θ_2 satisfies

$$n\left(\hat{\theta}_1 - \theta_1\right) = \frac{\frac{1}{n} \sum_{t=1}^{n} 1(v_t \leq \gamma) x_t u_t}{\frac{1}{n^2} \sum_{t=1}^{n} 1(v_t \leq \gamma) x_t^2} \Longrightarrow \frac{\int_0^1 W_x(r) dW_u(r,p)}{p \int_0^1 W_x^2(r) dr},$$

and

$$n\left(\hat{\theta}_2 - \theta_2\right) = \frac{\frac{1}{n} \sum_{t=1}^{n} 1(v_t > \gamma) x_t u_t}{\frac{1}{n^2} \sum_{t=1}^{n} 1(v_t > \gamma) x_t^2} \Longrightarrow \frac{\int_0^1 W_x(r) dW_u(r,q)}{q \int_0^1 W_x^2(r) dr}.$$

where $p = \Pr(v_t \leq \gamma)$ and $q = 1 - p$. Hence, $\hat{\theta}_1$ and $\hat{\theta}_2$ are both superconsistent in this case.

2.3.4 Testing for Co-summability

Given the asymptotic properties of the OLS estimator, OLS residuals can be used to construct a residual based test for co-summability. The following proposition formalizes this fact.

Proposition 11 *: Let $\hat{u}_t$ be the OLS residuals in equation (2.8):*

(i) Under Assumptions of Proposition 8

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} \hat{u}_t = O_p(1).$$

(ii) Under Assumptions of either Proposition 9 or 10

$$\frac{1}{n^{1/2+\delta_u}} \sum_{t=1}^{n} \hat{u}_t = O_p(1).$$

Remark: Under unbalancedness and spuriousness there is not a sequence of true errors u_t since no true model indeed exists. As it can be seen in case (i) of Proposition 11, in these situations, the OLS residuals will have the same order of summability than the endogeneous variable in the specified model. On the other hand, as showed in case (ii) of Proposition 11, the order of summability of the OLS residuals will be the order of summability of the true erros when there is co-summability, since an error or deviations from the equilibrium, u_t, exists in this case. These properties make it possible to consistently estimate the order of summability of the OLS residuals and, hence, to construct a test for co-summability using them. The following corollary formalizes this fact.

Corollary 1 *: Let $\hat{\beta}_{\hat{u}} = \left(1 + 2\hat{\delta}_{\hat{u}}\right)$ where $\hat{\delta}_{\hat{u}}$ is the order of summability estimator of the OLS residual in equation (2.8):*

(i) Under Assumptions of Proposition 8, if

$$\frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{[nr]} \hat{u}_t \Longrightarrow D_{\hat{u}}(r, \delta_y),$$

then

$$\log n \left(\hat{\beta}_{\hat{u}} - \beta_{\hat{u}}\right) \Longrightarrow \int_0^1 U_{\hat{u}}(r, \delta_y)\, dr,$$

where $U_{\hat{u}}(r, \delta_y) = \log\left[\left(r^{-1/2-\delta_y} D_{\hat{u}}(r, \delta_y)\right)^2\right]$.

(ii) Under Assumptions of either Proposition 9 or 10 if

$$\frac{1}{n^{1/2+\delta_u}} \sum_{t=1}^{[nr]} \hat{u}_t \Longrightarrow D_{\hat{u}}(r, \delta_u),$$

then

$$\log n \left(\hat{\beta}_{\hat{u}} - \beta_u\right) \Longrightarrow \int_0^1 U_{\hat{u}}(r, \delta_u)\, dr,$$

where $U_{\hat{u}}(r,\delta_u) = \log\left[\left(r^{-1/2-\delta_y} D_{\hat{u}}(r,\delta_u)\right)^2\right]$.

Given Proposition 11 and Corollary 1 a test for strong co-summability, $H_o : \delta_{\hat{u}} = 0$, can be easily constructed. First, estimate the order of summability of the residuals. Second, compute the corresponding subsampling confidence interval. Finally, check whether zero belongs to this interval. As before, the finite sample performance of the test will be studied via simulations. The data generating processes are those in Table 7. Again, performance has been measured by coverage probability of two-sided nominal 95% symmetric intervals for the null hypothesis. The experiment is based on 1000 replicas and three different sample sizes $n = \{100,\ 500,\ 1000\}$. A subsample size $b = \sqrt{n}$ has been chosen. Results are collected in Table 9.

Table 9: Coverage Probabilities. Testing for Strong Co-summability: $Ho : \delta_{\hat{u}} = 0$

S			n			**C**			n		
DGP	δ_y	δ_z	100	500	1000	DGP	δ_y	δ_z	100	500	1000
1	1/2	1/2	0.727	0.549	0.492	1	1/2	1/2	0.991	0.993	0.993
2	1/2	1/2	0.866	0.796	0.721	2	1/2	1/2	0.986	0.988	0.986
3	3/4	3/4	0.572	0.250	0.136	3	3/4	3/4	0.994	0.988	0.993
4	1	1	0.366	0.046	0.007	4	1	1	0.995	0.991	0.998
5	1	1	0.416	0.063	0.012	5	1	1	0.993	0.986	0.991
6	3/2	3/2	0.147	0.002	0.000	6	3/2	3/2	0.995	0.990	0.995
7	2	2	0.002	0.000	0.000	7	2	2	0.993	0.989	0.993
U1			n			**U2**			n		
DGP	δ_y	δ_z	100	500	1000	DGP	δ_y	δ_z	100	500	1000
1	1/2	0	0.931	0.867	0.823	1	0	1/2	0.986	0.988	0.986
2	1	0	0.155	0.008	0.001	2	0	1	0.995	0.991	0.998
3	3/2	0	0.105	0.009	0.000	3	0	3/2	0.995	0.990	0.995
4	2	0	0.087	0.008	0.001	4	0	2	0.993	0.989	0.993

S, C, $U1$, and $U2$ denote spuriousness, co-summability, unbalancedness of type 1, and unbalancedness of type 2, respectively.

Basically, under the null hypothesis of strong co-summability, C, the testing procedure shows a highly satisfactory performance in all cases even with very small samples sizes. As before, under S and U1 –alternative hypothesis– performance improves when we move away from the null hypothesis and increase the sample size. Notice that, by construction, in all the unbalanced of type 2 cases

$\delta_y = 0$, which means that the test is under the null hypothesis. Therefore, high coverage probabilities are found.

Remark: Let

$$y_t = m + \theta f(x_t) + u_t,$$

where m is an unknown constant term. In addition, let $\hat{m}$ and $\hat{\theta}$ the OLS estimator of m and θ, respectively. In this case,

$$\sum_{t=1}^{n} \hat{u}_t = 0,$$

which implies that $\hat{u}_t$ cannot be used to infer δ_u. The following pseudo residuals

$$\tilde{u}_t = y_t - \hat{\theta} f(x_t) = m + u_t - (\hat{\theta} - \theta) f(x_t),$$

could be used instead since

$$\sum_{t=1}^{n} \tilde{u}_t \neq 0.$$

The effect of m on the estimation of the order of summability of u_t can be tackled by an appropriate partial demeaning procedure –see section 1.4.4. in Chapter 1.

2.4 Empirical Application

2.4.1 Asymmetric preferences of central bankers

There is nowadays a great deal of consensus about the superiority of rules versus discretion in the practice of monetary policy. As pointed out by Taylor (1993) the advantage of rules over discretion is like the advantage of a cooperative over a non-cooperative solution in game theory. Optimal rules have been traditionally derived in a linear-quadratic framework in which policy makers have a quadratic objective function and operate in an economy that is described by a linear dynamic system –see for instance Svensson (1997). Linear Taylor rules are obtained in this framework when interest rates are taken to be the policy instrument implying that central banks adjust interest rates proportionally to inflation and output deviations from their targets. A traditional representative Taylor rule looks like

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_t + \theta_3 \tilde{y}_t, \tag{2.12}$$

where i_t denotes interest rates and $\tilde{\pi}_t$ and $\tilde{y}_t$ are deviations of inflation and output from their targets, respectively. Studying whether and how these rules are used when conducting monetary policy needs the estimation of the unknown parameters θ_0, θ_1, and θ_3. Using equation (2.12), or some slightly modified version of it, several authors have tried to quantify the parameters that define the practice of monetary policy in different countries –see, for instance, Clarida, Galí and Gertler (1998, 2000).

It is somehow surprising that little attention has been paid to the fact that variables involved in the Taylor rule are known to be highly persistent; something that should be taken into account when long time periods are analyzed. There are, however, several works that address this issue, for instance Siklos and Wohar (2005), Österholm (2005), or Christensen and Nielsen (2008). It seems to be a common feature of these studies the fact that traditional Taylor rules do not appear to be congruent with the data once persistence is taken into consideration –usually through integration and co-integration theory. This conclusion points to the possibility of an incorrect specification.

On the other hand, although in line with this conclusion, a stream of the literature has emphasized the hypothesis of asymmetric preferences of central bankers. This hypothesis is translated into non-linear Taylor rules. For instance, Clarida and Gertler (1997) consider a threshold type of Taylor rule in which the reaction of the monetary authority is different when inflation or output deviates from above than from below the target. Specifically,

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_{t+k} 1\left(v_t > 0\right) + \theta_2 \tilde{\pi}_{t+k} 1\left(v_t \leq 0\right) + \theta_3 \tilde{y}_t 1\left(v_t > 0\right) + \theta_4 \tilde{y}_t 1\left(v_t \leq 0\right), \tag{2.13}$$

where $\tilde{\pi}_{t+k}$ is deviations of the rate of inflation between periods t and $t+k$, and v_t can be either $\tilde{\pi}_{t+k}$ or $\tilde{y}_t$. Alternatively, Dolado, María-Dolores and Naveira (2005) derive a non-linear optimal rule when non-linearities in the Phillips curve are allowed. The main prediction of this model is that the Taylor rule should contain the interaction between inflation and output gaps, that is

$$i_t = \theta_0 + \theta_1 \tilde{\pi}_{t+k} + \theta_2 \tilde{y}_t + \theta_3 \tilde{\pi}_{t+k} \tilde{y}_t. \tag{2.14}$$

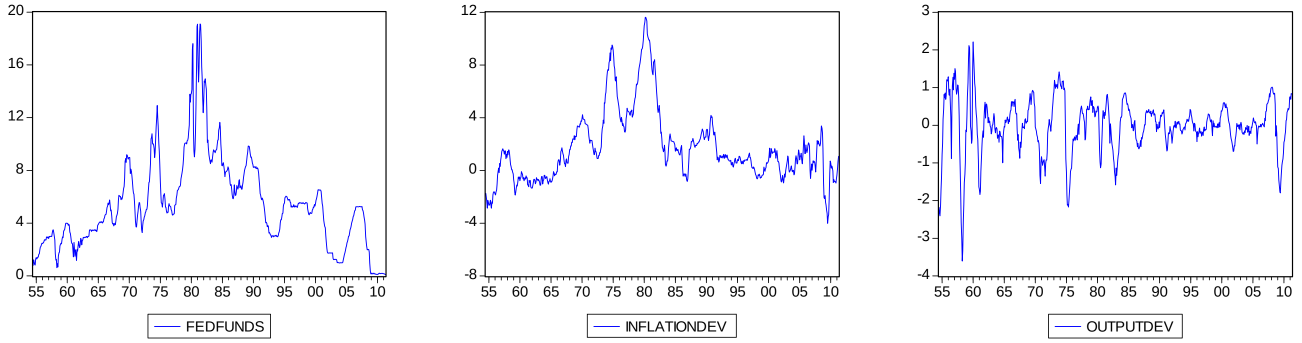
Notice that if i_t, $\tilde{\pi}_{t+k}$, or $\tilde{y}_t$ are highly persistent, the non-linear nature of these two specifications invalidate the use of standard co-integration theory to analyze the relevance of these models. Nevertheless, co-summability can be used to test for these two hypothesis given its generality when allowing for persistence and non-linearities at the same time. The linearity in parameters of both equations makes suitable the application of the above developed tools to test for co-summability.

To this end, we use US monthly time series covering the period 1954:07-2011:04, which are obtained from the Federal Reserve Bank of St. Louis. Specifically, we use (i) Federal Funds rate as interest rates, (ii) annual ($t/t-12$ basis; $k=12$) percentage rate in the CPI for inflation, (iii) (logged) Industrial Production Index for output. Following the usual practice in the literature, to measure the output gap, we detrend (logged) industrial production using the HP filter with a coefficient of 14.800. For the inflation target, we use a fixed 2% level. Figures in Table 10 show the temporal evolution of these three measures –i_t, $\tilde{\pi}_{t+k}$, and $\tilde{y}_t$.

On the other hand, Table 11 reports the estimated orders of summability of all the variables contained in equations (2.13) and (2.14) as well as their corresponding subsampling confidence

Table 10: Optimal Rules of Monetary Policy

Variables: Taylor Rule



interval[9]. All the variables have been partially demeaned to compute their orders of summability. Moreover, to control for a possible constant term in regression model (2.9) the first observation is substracted –see Section 1.4.2 in Chapter 1 for details.

Results in Table 11 indicate that interest rates, i_t, and inflation gap, $\tilde{\pi}_{t+k}$, have a similar order of summability around 0.8 while the estimated order of summability for the output gap, $\tilde{y}_t$, is around 0.5. It is worth to emphasize that zero does not belong to any of the subsampling confidence intervals for these three time series. Hence, persistence has to be properly addressed when using this database. With respect the non-linear variables, different results are found. The cross-product $\tilde{\pi}_{t+k}\tilde{y}_t$ presents a lower order of summability estimate, while the threshold transformations tend keep the the order of summability. Misspecifications of traditional Taylor rules could be explained if these processes are relevant for the Fed when conducting monetary policy.

Following the steps of the proposed empirical strategy to test for co-summability, next, the balancedness of equations (2.13) and (2.14) is analyzed[10]. As it can be seen in Table 12, the null hypothesis of balancedness, $H_o : \delta_y = \delta_z$, is not rejected in all the cases but for $\tilde{y}_t 1\left(\tilde{\pi}_{t+k} > 0\right)$. Notice that even in this case, the postulated model (2.13) would be still balanced since $\delta_y - \delta_z > 0$ for $\tilde{y}_t 1\left(\tilde{\pi}_{t+k} > 0\right)$ but $\delta_y - \delta_z = 0$ for all the other explanatory variables in the model. Given these

[9]Notice that threshold variables can start taking the value zero; something that is not allowed in the estimation procedure by Assumption 1. In these cases we have just eliminated the observations corresponding to the first set of zeros.

[10]The test for balancedness is based on the difference $\hat{\delta}_y - \hat{\delta}_z$, where $\hat{\delta}_y$ and $\hat{\delta}_z$ are the orders of summability of the endogenous variable and some of the explanatory variables, respectively. It is important to mention that $\hat{\delta}_y$ and $\hat{\delta}_z$ are obtained directly from regressions (2.9) and (2.10) without substracting the first observation. Monte Carlo experiments showed that subsampling confidence intervals for balancedness performs much better in this way.

Table 11: Order of Summability: Estimation and Inference

Variables	$\hat{\delta}^*$	I_L	I_U
i_t	0.820	0.419	1.221
$\tilde{\pi}_{t+k}$	0.869	0.404	1.333
$\tilde{y}_t$	0.495	0.059	0.930
$\tilde{\pi}_{t+k}\tilde{y}_t$	0.200	-0.363	0.764
$\tilde{\pi}_{t+k}1\left(\tilde{\pi}_{t+k} > 0\right)$	0.825	0.464	1.186
$\tilde{\pi}_{t+k}1\left(\tilde{\pi}_{t+k} \leq 0\right)$	0.701	0.295	1.106
$\tilde{y}_t1\left(\tilde{\pi}_{t+k} > 0\right)$	0.150	-0.495	0.797
$\tilde{y}_t1\left(\tilde{\pi}_{t+k} \leq 0\right)$	0.412	-0.225	1.051
$\tilde{\pi}_{t+k}1\left(\tilde{y}_t > 0\right)$	0.811	0.301	1.321
$\tilde{\pi}_{t+k}1\left(\tilde{y}_t \leq 0\right)$	0.733	0.239	1.228
$\tilde{y}_t1\left(\tilde{y}_t > 0\right)$	0.496	0.137	0.855
$\tilde{y}_t1\left(\tilde{y}_t \leq 0\right)$	0.628	0.191	1.065

$\hat{\delta}^*$ denotes the estimated order of summability from a regression that substracts the first observation. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals. All the variables have been partially demeaned.

results, Step 2 in the proposed empirical strategy –testing for co-summability– should be carried out.

The test for co-summability is designed from a residual based statistic. Therefore, an estimate of the parameters of the model is needed. Table 13 collects these estimates jointly with the results of the test for co-summability associated to each regression. Some aspects are worth to be emphasized. First, as it can be seen in column 2 of Table 13, the traditional Taylor rule does not specify a co-summable relationship. Second, focusing on the non-linear specifications, it can be seen that only a threshold type of Taylor rule in which the Federal Reserve reacts asymmetrically to output deviations –column 5 of Table 13– specifies a co-summable relationship. This is totally in line with the view that, contrary to the Bundesbank, the Federal Reserve reacts more to output than inflation movements. Finally, it is remarkable the difference between the parameters associated to $\tilde{y}_t1\left(\tilde{y}_t > 0\right)$ and $\tilde{y}_t1\left(\tilde{y}_t \leq 0\right)$. This fact clearly favors the stream of the literature that highlights the greater aversion to recessions than to expansions of the monetary authorities in the US.

Table 12: Testing for Balancedness

Balancedness	$\hat{\delta}_y - \hat{\delta}_z$	I_L	I_U
$\tilde{\pi}_{t+k}$	0.034	-0.430	0.500
$\tilde{y}_t$	0.493	-0.114	1.101
$\tilde{\pi}_{t+k}\tilde{y}_t$	0.488	-0.062	1.038
$\tilde{\pi}_{t+k}1\left(\tilde{\pi}_{t+k} > 0\right)$	0.084	-0.317	0.486
$\tilde{\pi}_{t+k}1\left(\tilde{\pi}_{t+k} \leq 0\right)$	0.202	-0.317	0.722
$\tilde{y}_t 1\left(\tilde{\pi}_{t+k} > 0\right)$	0.637	0.067	1.207
$\tilde{y}_t 1\left(\tilde{\pi}_{t+k} \leq 0\right)$	0.576	-0.472	1.624
$\tilde{\pi}_{t+k}1\left(\tilde{y}_t > 0\right)$	0.072	-0.363	0.508
$\tilde{\pi}_{t+k}1\left(\tilde{y}_t \leq 0\right)$	0.170	-0.292	0.632
$\tilde{y}_t 1\left(\tilde{y}_t > 0\right)$	0.382	-0.237	1.002
$\tilde{y}_t 1\left(\tilde{y}_t \leq 0\right)$	0.360	-0.208	0.929

$\hat{\delta}_y$ and $\hat{\delta}_z$ denote the estimated order of summability of the endogenous variable and the specified explanatory variable, respectively. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals. All the variables have been partially demeaned

2.4.2 Environmental Kuznets Curve

The Environmental Kuznets Curve –EKC– suggests an inverted U-shaped relationship between pollution and economic development. The argument is as follows. Agents living in poor economies are more concerned with jobs and income than clean air and water, consequently environmental regulation is weak. As economies get richer, agents value more the environment, production becomes cleaner, and regulatory institutions are more efficient.

This hypothesis has been controversial, prompting confronted views from researchers and policymakers. The literature –see Grossmann and Krueger (1995) or Brock and Taylor (2005)– identifies, mainly, three different channels linking pollution and economic activity: scale, composition, and technique effects. Ceteris paribus (i) emissions rise and fall in proportion to the scale of the economic activity as measured by real GDP; (ii) emissions fall via the pure composition effect if an economy moves towards producing goods that are cleaner on average than those they produced before; (iii) emissions can fall when the techniques of production become cleaner. The EKC hypothesis will depend on the relative importance of these three effects. To identify them, a structural modelling must be carefully undertaken. Nevertheless, the empirical literature on the EKC has always

Table 13: Testing for Co-summability

Taylor Rules	i_t	i_t	i_t	i_t
1	3.806	3.772	3.833	4.083
$\tilde{\pi}_{t+k}$	0.938	0.941		
$\tilde{y}_t$	0.753	0.458		
$\tilde{\pi}_{t+k}\tilde{y}_t$		0.173		
$\tilde{\pi}_{t+k}1(\tilde{\pi}_{t+k} > 0)$			0.930	
$\tilde{\pi}_{t+k}1(\tilde{\pi}_{t+k} \leq 0)$			1.035	
$\tilde{y}_t1(\tilde{\pi}_{t+k} > 0)$			0.996	
$\tilde{y}_t1(\tilde{\pi}_{t+k} \leq 0)$			0.247	
$\tilde{\pi}_{t+k}1(\tilde{y}_t > 0)$				1.048
$\tilde{\pi}_{t+k}1(\tilde{y}_t \leq 0)$				0.787
$\tilde{y}_t1(\tilde{y}_t > 0)$				-0.217
$\tilde{y}_t1(\tilde{y}_t \leq 0)$				1.003
$\hat{\delta}^*_{\hat{u}}$	0.449	0.493	0.443	0.371
I_L	0.040	0.073	0.022	-0.033
I_U	0.859	0.913	0.865	0.777

$\hat{\delta}^*_u$ denote the estimated order of summability of the residuals in each regression. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals.

used a reduced form approach. Typically, polynomial relationships between pollution and income has been considered, that is,

$$p_t = \theta_0 + \theta_1 y_t + \theta_2 y_t^2 + ... + \theta_k y_t^k, \tag{2.15}$$

where p_t is a measure of pollution and y_t is a measure of income. Several empirical issues arise in this setup. A first issue is concerned with the measures chosen for p_t and y_t. While GDP has been usually used as a measure of income, y_t, many measures of pollutants have been used. Commonly used measures for p_t are CO_2, NO_x, and SO_2. Empirical evidence is mixed for different pollutants. A second issue relates to the curvature of the EKC. There seems not to be a clear agreement about the order of the polynomial to be used. Grossman and Krueger (1995) used a cubic specification, while Holtz-Eakin and Selden (1995) preferred the quadratic one. Other authors tend to compare both specifications in practice. A third empirical ambiguity arise since p_t and y_t are sometimes treated in levels (Grossman and Krueger, 1995), others in logarithms (Hong and Wagner, 2008), or both cases are compared (Holtz-Eakin and Selden, 1995). Finally, it is surprising that only few authors have taken into consideration persistence of the variables involved in the EKC. Some exceptions being Perman and Stern (2003), Hong and Wagner (2008) and Jalil and Mahmud (2009). When persistence is taken into consideration, the empirical evidence on the EKC is confusedly mixed.

Next, with illustrative purposes, we apply co-summability theory to this hypothesis trying to elucidate some empirical features on the EKC. We use annual GDP and CO_2 emissions in the US during the period 1870-2007. GDP and population are taken from Angus Maddison and CO_2 emissions from the Carbon Dioxide Information Analysis Center. Table 14 shows the evolution of GDP and CO_2 emissions per capita, both in levels –$co2$, gdp– and logarithms –$lco2$, $lgdp$.

Table 14: Environmental Kuznets Curve Hypothesis

Variables: Environmental Kuznets Curve

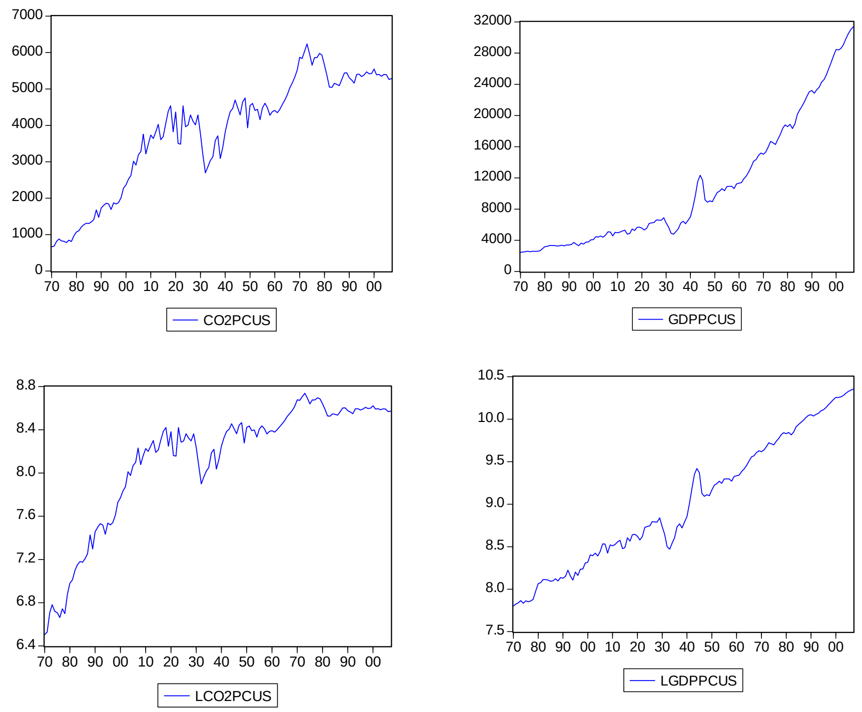


Table 15 reports the estimated orders of summability of all the variables contained in equation (2.15) for $k = 4$. The corresponding subsampling confidence intervals are provided as well. As expected, the order of summability of GDP per capita increases as successive powers are taken. There is, however, a non-negligible difference between the orders of summability of gdp^k and $lgdp^k$. In general, these results show that persistence must be taken into account in this non-linear framework. Moreover, given the linearity in parameters of the empirical reduced forms of the EKC, the tools developed above are suitable to be applied.

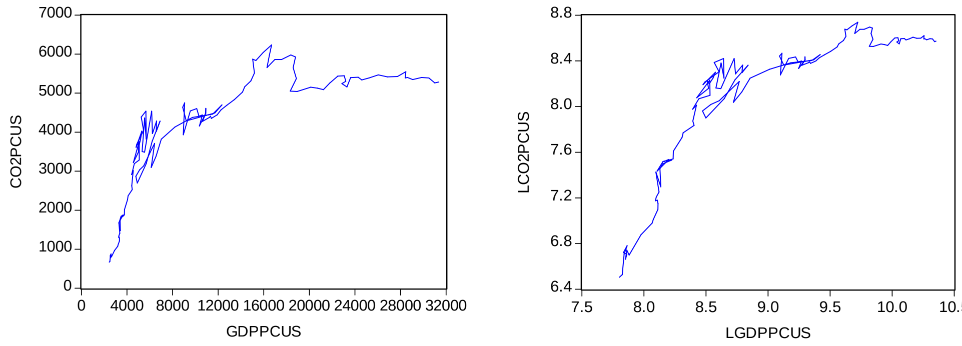
Table 15: Order of Summability: Estimation and Inference

Variables	$\hat{\delta}^*$	I_L	I_U
$co2$	0.985	0.405	1.566
gdp	1.448	0.662	2.234
gdp^2	1.796	0.796	2.797
gdp^3	2.157	0.898	3.416
gdp^4	2.516	1.101	3.932
$lco2$	0.807	0.288	1.325
$lgdp$	1.030	0.347	1.713
$lgdp^2$	1.085	0.381	1.789
$lgdp^3$	1.132	0.391	1.874
$lgdp^4$	1.196	0.408	1.984

$\hat{\delta}^*$ denotes the estimated order of summability from a regression that substracts the first observation. I_L^* and I_U^* denote the lower and upper bounds of the corresponding subsampling intervals. All the variables have been partially detrended.

Before proceeding to the quantitative multivariate analysis –balancedness and co-summability–, it seems interesting to plot the relationship between GDP and CO_2 emissions in a graph. Table 16 contains these figures for both the levels and the logs of the variables in the EKC. Although, it seems there is a diminishing marginal propensity to emit in the US, the postulated inverted U-shape should be more carefully and formally analyzed.

Table 16: Environmental Kuznets Curve Hypothesis



Results of testing for balancedness are reported in Table 17. Notice that when the variables are in levels, the quadratic specification is unbalanced. Cubic or higher order polynomials will not generate a balanced model neither. However, when logarithms are taken, quadratic and cubic specifications become balanced. Notice that in this case a linear relationship between $lco2$ and $lgdp$ is unbalanced. The same occurs with polynomials of degree $k \geq 4$. These results reveals a peculiar empirical fact regarding the EKC in the US. The levels of GDP and CO_2 emissions do not describe an inverted U-shape, at least with a polynomial specification. Other specifications should not be ruled out.

Table 17: Testing for Balancedness

	Balancedness	$\hat{\delta}_y - \hat{\delta}_z$	I_L	I_U
$co2$				
	gdp	-0.104	-0.995	0.787
	gdp^2	-2.494	-4.809	-0.179
$lco2$				
	$lgdp$	0.530	0.054	1.007
	$lgdp^2$	-0.203	-0.956	0.550
	$lgdp^3$	-0.860	-1.935	0.215
	$lgdp^4$	-1.517	-2.916	-0.117

$\hat{\delta}_y$ and $\hat{\delta}_z$ denote the estimated order of summability of the endogenous variable and the specified explanatory variable, respectively. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals. All the variables have been partially detrended.

Despite the results found when testing for balancedness, it seems appropriate to report results of co-summability analysis for the levels of the variables. These results are collected in Table 18.

As it can be seen, both linear and quadratic specifications are studied. Notice that neither the linear nor the quadratic cases specify a co-summable relationship. This conclusion does not change when a trend is introduced in the estimated regression –a usual practie in this literature. These results are conclusive and go against the polynomial specification for the EKC hypothesis in the US. We do not interpret this results as an strict refutation of the EKC. At most, we can say that the polynomial reduced form is misspecified when using the levels of $co2$ and gdp.

Table 18: Testing for Co-summability

EKC	$co2$	$co2$	$co2$	$co2$
1	1771.517	1963.832	130.940	235.162
t		-2.706		-1.299
gdp	0.172	0.177	0.569	0.569
gdp^2			-1.389e-005	-1.379e-005
$\hat{\delta}^*_{\hat{u}}$	1.656	1.604	1.206	1.284
I_L	0.846	0.850	0.509	0.527
I_U	2.466	2.358	1.903	2.040

$\hat{\delta}^*_u$ denote the estimated order of summability of the residuals in each regression. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals.

The previous pessimistic conclusion using levels is reverted to optimism when the logarithms of GDP and CO_2 are studied. In this case, when $k = 1$, the relationship is not co-summable. This fact was expected given the results obtained when testing for balancedness in this case. However, quadratic as well as a cubic specification appears to be co-summable relationships. Notice that this conclusion is independent of whether a linear trend is introduced in the regression function or not. In addition, it is worth to emphasize that the sign of the parameters totally agree with the predictions of the EKC hypothesis. Logged emissions in the US increased with logged GDP until a certain level; once it was reached, emissions decreased.

Given the properties of the OLS estimator under unbalancedness, spuriousness and co-summability, we strongly recommend taking logarithms when empirically studying polynomial reduced forms of

Table 19: Testing for Co-summability

EKC	$lco2$	$lco2$	$lco2$	$lco2$	$lco2$	$lco2$
1	1.527	1.461	-46.944	-47.130	-280.561	-280.684
t		-0.001		0.0001		-2.391e-005
$lgdp$	0.727	0.745	11.558	11.599	89.890	89.935
$lgdp^2$			-0.600	-0.602	-9.319	-9.324
$lgdp^3$					0.322	0.322
$\hat{\delta}^*_{\hat{u}}$	1.373	1.404	0.413	0.440	0.116	0.127
I_L	0.654	0.628	-0.078	-0.029	-0.479	-0.458
I_U	2.091	2.180	0.904	0.910	0.712	0.713

$\hat{\delta}^*_u$ denote the estimated order of summability of the residuals in each regression. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals.

the EKC, at least in the US.

2.5 Concluding Remarks

Co-integration theory is not designed to deal with situations in which non-linearities and persistence occur at the same time. There is a clear need for theoretically valid and empirically useful concepts that generalize those of integration and co-integration to non-linear environments.

The order of summability concept has made possible to define nonlinear long run relationships between persistent processes under exactly the same logic as the one of co-integration theory. It has easily allowed (i) to define balancedness of a postulated model –a necessary condition for a correct specification; and (ii) to define non-linear long run relationships by means of the concept of co-summability –a direct extension of co-integration valid for non-linear equilibria. These two pieces are relevant for both econometricians and economic theorists: for the former when specifying, estimating, and testing econometric models; for the latter when choosing functional forms to construct their theories.

In a non-linear in variables but linear in parameters framework, the statistical tools to deal with univariate summable processes can be used to design tests for balancedness and co-summability. Both tests contribute to improve upon the statistical treatment of these type of non-linear regression models making them fully operative in practice. Indeed, the concepts and tools developed in this chapter offer a new set of econometric possibilities in the study of non-linear long run relationships generalizing co-integration theory.

2.6 Appendix

Proof of Proposition 7: It was shown in Chapter 1 that if

$$S_{y[nr]}(\delta) = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{[nr]} y_t \Longrightarrow D_y(r, \delta_y),$$

then

$$\log n \left(\hat{\beta}_y - \beta_y\right) = \log n \frac{\sum_{k=1}^{n} U_{yk} \log k}{\sum_{k=1}^{n} \log^2 k} \Longrightarrow \int_0^1 \log\left(D_y^2(r, \delta_y)\right) dr.$$

Now,

$$\hat{\beta}_n = \hat{\beta}_y - \hat{\beta}_z = \beta_y + \frac{\sum_{k=1}^{n} U_{yk} \log k}{\sum_{k=1}^{n} \log^2 k} - \beta_z - \frac{\sum_{k=1}^{n} U_{zk} \log k}{\sum_{k=1}^{n} \log^2 k}$$

or equivalently

$$\hat{\beta}_n - \beta = \left(\hat{\beta}_y - \beta_y\right) - \left(\hat{\beta}_z - \beta_z\right) = \frac{\sum_{k=1}^{n} (U_{yk} - U_{zk}) \log k}{\sum_{k=1}^{n} \log^2 k}.$$

Therefore,

$$\log n \left(\hat{\beta}_n - \beta\right) = \log n \frac{\sum_{k=1}^{n} (U_{yk} - U_{zk}) \log k}{\sum_{k=1}^{n} \log^2 k} \Longrightarrow \int_0^1 (U_y(r, \delta_y) - U_z(r, \delta_z))\, dr.$$

Q.E.D.

Proof of Proposition 8*:* The rescaled OLS estimator can be written as

$$n^{\alpha_z-\alpha_y}\hat{\theta}_n = \frac{\frac{1}{n}\sum_{t=1}^{n}\frac{y_t}{n^{\alpha_y}}\frac{f(x_t)}{n^{\alpha_z}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{f^2(x_t)}{n^{2\alpha_z}}} = \frac{\int_0^1 \frac{y_{[nr]}}{n^{\alpha_y}}\frac{f\left(x_{[nr]}\right)}{n^{\alpha_z}}dr}{\int_0^1 \frac{f^2\left(x_{[nr]}\right)}{n^{\alpha_z}}dr}.$$

Now, under Assumption 2 the CMT can be used to get the stated result

$$n^{\alpha_z-\alpha_y}\hat{\theta}_n \Longrightarrow \frac{\int_0^1 D_z\left(r,\alpha_z\right)D_y\left(r,\alpha_y\right)dr}{\int_0^1 D_z^2\left(r,\alpha_z\right)dr}.$$

Q.E.D.

Proof of Proposition 9: The OLS estimator in terms of u_t and $f\left(x_t\right)$,

$$\hat{\theta}_n = \theta + \frac{\sum_{t=1}^{n}u_t f\left(x_t\right)}{\sum_{t=1}^{n}f^2\left(x_t\right)},$$

can be expressed and rewritten as

$$n^{\alpha_z-\alpha_u}\left(\hat{\theta}_n-\theta\right) = \frac{\frac{1}{n}\sum_{t=1}^{n}\frac{u_t}{n^{\alpha_u}}\frac{f(x_t)}{n^{\alpha_z}}}{\frac{1}{n}\sum_{t=1}^{n}\frac{f^2(x_t)}{n^{2\alpha_z}}} = \frac{\int_0^1 \frac{u_{[nr]}}{n^{\alpha_y}}\frac{f\left(x_{[nr]}\right)}{n^{\alpha_z}}dr}{\int_0^1 \frac{f^2\left(x_{[nr]}\right)}{n^{\alpha_z}}dr}.$$

Hence, under Assumptions 3, applying the CMT

$$n^{\alpha_z-\alpha_u}\left(\hat{\theta}_n-\theta\right) \Longrightarrow \frac{\int_0^1 D_z\left(r,\alpha_z\right)D_u\left(r,\alpha_u\right)dr}{\int_0^1 D_z^2\left(r,\alpha_z\right)dr}.$$

Proof of Proposition 10: The denominator of the OLS estimator

$$n^{\alpha_z}\left(\hat{\theta}_n-\theta\right) = \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{f(x_t)}{n^{\alpha_z}}u_t}{\frac{1}{n}\sum_{t=1}^{n}\frac{f^2(x_t)}{n^{2\alpha_z}}}.$$

has been previously studied. Next, the numerator will be analyzed. Let

$$v_{nt} = \frac{1}{\sqrt{n}}\sum_{i=1}^{t}u_i,$$

such that

$$v_{nt} - v_{n,t-1} = \frac{u_t}{\sqrt{n}}.$$

and its D-space analog

$$dv_n\left(r\right) = \frac{u_{[nr]}}{\sqrt{n}}.$$

Next, we concentrate on

$$\sum_{t=1}^{[nr]} \frac{f\left(x_{[nr]}\right)}{n^{\alpha_z}} \frac{u_{[nr]}}{\sqrt{n}} = \int_0^r z_n\left(r\right) dv_n\left(r\right).$$

Following Hansen (1992), define

$$\epsilon_i = \sum_{k=0}^{\infty} \left(E_i u_{i+k} - E_{i-1} u_{i+k}\right), \qquad w_i = \sum_{k=1}^{\infty} E_i u_{i+k},$$

such that

$$u_i = \epsilon_i + w_{i-1} - w_i, \qquad E_{i-1}\epsilon_i = 0.$$

In this scenario, a martingale difference approximation can be used, that is,

$$\int_0^r z_n\left(r\right) dv_n\left(r\right) = \int_0^r z_n\left(r\right) dY_n\left(r\right) + \Lambda_n^*\left(r\right),$$

where $Y_n\left(r\right) = Y_{n[nr]}$, $Y_{nt} = Y_t/\sqrt{n}$, $Y_t = \sum_i^t \epsilon_i$, and

$$\Lambda_{nt}^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{t} \left(z_{ni} - z_{n,i-1}\right) w_i - \frac{1}{\sqrt{n}} z_{nt} w_{t+1}.$$

Let $\epsilon_{ni} = \epsilon_i/\sqrt{n}$. To apply Theorem 3.1. in Hansen (1992), that is, to get

$$\int_0^r z_n\left(r\right) dY_n\left(r\right) \Longrightarrow \int_0^r D_z\left(r\right) dD_u\left(r\right),$$

it must be showed that:

(i)

$$\sum_{i=1}^{n} E\epsilon_{ni}^2 < \infty,$$

(ii)

$$\left(Y_n\left(r\right) - v_n\left(r\right)\right) \xrightarrow{p} 0.$$

With respect (i), note that

$$\begin{aligned}
\sum_{i=1}^{n} E\epsilon_{ni}^2 \leq \sup_{i\leq n} E\epsilon_i^2 &= \left(\sup_{i\leq n}\left|u_i - w_{i-1} + w_i\right|\right)^2 \\
&\leq \left(\sup_{i\leq n}\left|u_i\right| + \sup_{i\leq n}\sum_{k=1}^{\infty}\left|\left(E_i u_{i+k} - E_{i-1}u_{i-1+k}\right)\right|\right)^2.
\end{aligned}$$

On one hand, by condition (a)

$$\sup_{i\leq n}\left|u_i\right| < \infty.$$

On the other hand,

$$\begin{aligned}
\sup_{i\leq n}\sum_{k=1}^{\infty}\left|\left(E_i u_{i+k} - E_{i-1}u_{i-1+k}\right)\right| &\leq \sup_{i\leq n}\sum_{k=1}^{\infty}\left(\left|E_i u_{i+k}\right| + \left|E_{i-1}u_{i-1+k}\right|\right) \\
&= \sup_{i\leq n}\sum_{k=1}^{\infty}\left|E_i u_{i+k}\right| + \sup_{i\leq n}\sum_{k=1}^{\infty}\left|E_{i-1}u_{i-1+k}\right| < \infty,
\end{aligned}$$

by condition (b).

With respect (ii), note that

$$\sup_{i\leq n}|Y_{nt}-v_{nt}|\leq 2\frac{1}{\sqrt{n}}\sup_{i\leq n}|w_t|\xrightarrow{p}0,$$

by condition (b).

It remains to analyze

$$\Lambda_{nt}^{*}=\frac{1}{\sqrt{n}}\sum_{i=1}^{t}\left(z_{ni}-z_{n,i-1}\right)w_i-\frac{1}{\sqrt{n}}z_{nt}w_{t+1}.$$

First, consider

$$\sup_{t\leq n}\frac{1}{\sqrt{n}}\left|z_{nt}w_{t+1}\right|\leq\sup_{t\leq n}|z_{nt}|\sup_{i\leq n}\frac{1}{\sqrt{n}}\left|w_{t+1}\right|.$$

By the assumptions on z_{nt}

$$\sup_{t\leq n}|z_{nt}|=O_p\left(1\right),$$

and by condition (b)

$$\frac{1}{\sqrt{n}}\sup_{t\leq n}|w_t|\xrightarrow{p}0.$$

Therefore,

$$\frac{1}{\sqrt{n}}z_{nt}w_{t+1}\xrightarrow{p}0.$$

With respect

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{t}\left(z_{ni}-z_{n,i-1}\right)w_i=\frac{1}{\sqrt{n}}\sum_{i=1}^{t}\frac{f\left(x_i\right)-f\left(x_{i-1}\right)}{n^{\alpha_z}}\sum_{k=1}^{\infty}E_iu_{i+k},$$

note that

$$\begin{aligned}\frac{1}{n^{\alpha_z}}\sum_{i=1}^{t}\left[f\left(x_i\right)-f\left(x_{i-1}\right)\right]\frac{1}{\sqrt{n}}\sum_{k=1}^{\infty}E_iu_{i+k}&\leq\frac{1}{\sqrt{n}}\sup_{i\leq t}\sum_{k=1}^{\infty}\left|E_iu_{i+k}\right|\frac{1}{n^{\alpha_z}}\sum_{i=1}^{t}\left[f\left(x_i\right)-f\left(x_{i-1}\right)\right]\\&=\frac{1}{\sqrt{n}}\sup_{t\leq n}\sum_{k=1}^{\infty}\left|E_tu_{t+k}\right|\left(\frac{f\left(x_t\right)}{n^{\alpha_z}}-\frac{f\left(x_0\right)}{n^{\alpha_z}}\right)\\&=o_p\left(1\right)O_p\left(1\right)+o_p\left(1\right)\\&=o_p\left(1\right).\end{aligned}$$

Therefore,

$$\frac{1}{n^{1/2+\alpha_z}}\sum_{t=1}^{n}f\left(x_t\right)u_t\Longrightarrow\int_0^1D_z\left(r\right)dD_u\left(r\right).$$

Finally,

$$n^{\alpha_z}\left(\hat{\theta}_n-\theta\right)\Longrightarrow\frac{\int_0^1D_z\left(r,\alpha_z\right)dW\left(r\right)}{\int_0^1D_z^2\left(r,\alpha_z\right)dr},$$

as stated. **Q.E.D.**

Proof of Proposition 11:

(i) Under Assumptions of Proposition 8

$$\frac{1}{n^{1/2+\delta_y}}\sum_{t=1}^{n}\hat{u}_t=\frac{1}{n^{1/2+\delta_y}}\sum_{t=1}^{n}y_t-n^{\delta_z-\delta_y}\hat{\theta}_n\frac{1}{n^{1/2+\delta_z}}\sum_{t=1}^{n}f\left(x_t\right)=O_p\left(1\right).$$

(ii) Under Assumptions of Proposition 9 or 10

$$\frac{1}{n^{1/2+\delta_u}}\sum_{t=1}^{n}\hat{u}_t=\frac{1}{n^{1/2+\delta_u}}\sum_{t=1}^{n}u_t-n^{\delta_z-\delta_u}\left(\hat{\theta}_n-\theta\right)\frac{1}{n^{1/2+\delta_z}}\sum_{t=1}^{n}f\left(x_t\right)=O_p\left(1\right).$$

Q.E.D.

Proof of Corollary 1: The proof follows directly from Proposition 11 above and Proposition 4 in Chapter 1. **Q.E.D.**

Chapter 3

The Threshold Impatient Investor Case

Abstract: This chapter studies the implications of changes in impatience of investors on asset pricing. In a standard consumption based capital asset pricing model with risk neutral agents, assets are priced via a present value formula with a constant discount factor determined by the constant level of patience of the investors. However, changes in impatience along their lifetimes would imply a changing discount factor to price assets in the economy. To capture such changes and analyze its impact on asset pricing, we model investors with a level of patience that randomly depends on whether the future is perceived as a good or bad state of the nature. The model produces an optimal threshold discounting process for valuing assets, which we call the *threshold present value model*. From this optimal threshold pricing formula, it is shown that prices and dividends share a theoretical threshold long run equilibrium relationship.

Using US stock market data, the threshold present value model is tested empirically. Since persistence and non-linearities characterize the empirical analysis co-summability theory is used.

3.1 Introduction

In everyday language, the term *patience* has a number of different meanings. We usually understand it as calm and tranquility when one waits, as the capacity to tolerate a suffering or something hard or annoying, or as tranquility to do minute or difficult tasks. In economics, patience is modelled as a subjective discount factor. It should be interpreted as the weight that future consequences have in today decisions, as reflected in savings rates, for example. This utility weight can be interpreted as a reduced-form representation of a non-cognitive skill, such as the ability to exert self-control –in environments with temptation–, or the ability to imagine the future vividly.

The most standard treatment of impatience in the economic literature assumes that it is constant along the lifetime of the agents. However, the discounted utility model of intertemporal choice is contradicted by a relatively large body of empirical and experimental evidence. Moreover, taking constant the rate of time preference there exists little discussion of what determines its level. The response of the literature to these weaknesses has been to model patience, the discount factor, in several different ways.

One of the earliest approaches, trying to overcome the first weakness, is the hyperbolic discounting. Research on animal and human behavior has led psychologists and economists to conclude –see Ainslie (1992), Ainslie and Haslam (1992), and Loewenstein and Prelec (1992)– that discount functions are generalized hyperbolas. Such discount functions imply a monotonically falling discount rate. This discount structure sets up a conflict between today preferences and the preferences which will be held in the future, implying that preferences are dynamically inconsistent. From today's perspective, the discount rate between two far off periods, t and $t+1$, is a long-term low discount rate. However, from the time t perspective, the discount rate between t and $t+1$ is a short-term high discount rate.

To overcome the second weakness, that is, to introduce more discussion about the determination of the rate of time preference, Becker and Mulligan (1997) constructed a model of patience formation. It combines the classical economists insights with a particular view of what it means to be rational, a conception of rationality that is consistent with many kinds of human frailties, including defective recognition of future utilities. Rational persons may spend resources in the attempt to overcome their frailties. This simple idea provides the point of departure for the analysis in Becker and Mulligan (1997) to endogenize time preference. Hence, unlike the usual neoclassical approach, they do not assume that the discount factor is a fixed parameter, but rather that it is adjusted according to the propinquity of future pleasures. Specifically, they assume that people have the option to put effort to increase their appreciation of the future. They model such effort by allowing a consumer to make future pleasures less remote by spending resources on imagining them.

Changes in impatience of agents may potentially have several economic implications since time preference play a fundamental role in theories of savings and investment, economic growth, interest rate or asset pricing determination, addiction, and many other issues that are getting increasing attention from economists. We concentrate here on the implications for asset pricing.

The theoretical benchmark in which we base our analysis is the standard intertemporal consumption based capital asset pricing model –denoted C-CAPM. Specifically, in such a framework, the investor maximizes expected utility which depends only on current and future consumption –see Lucas (1978). Financial assets play a role in this model in that they help to smooth consumption over time. Securities are held to transfer purchasing power from one period to another. The optimality condition in the standard C-CAPM is a "*standard present value statement*". It is worth to emphasize that the standard present value formula can be derived too from the definition of expected returns assuming they are constant –as patience of standard risk neutral investors.

In our analysis, with the C-CAPM as a benchmark, the impatience of an investor changes according to a threshold variable. The threshold structure of the subjective discount factor allows for time variation, which depends on whether the agents perceive the future as a good or bad state of the nature. This perception is modelled as a random process instead of the endogenous determination in Becker and Mulligan (1997). Uncertainty in the economy makes agents not able to fully determine their level of patience. Random elements will affect this level each period. In this scenario, the optimality condition is again a present value statement, but of a threshold type. We call this solution the "*threshold present value model*". The same optimality condition is derived from the definition of expected returns, now assuming they follow a threshold process –as the patience of risk neutral threshold impatient investors.

Following Campbell and Shiller (1987), a threshold long run equilibrium relationship between prices and dividends is found. Specifically, assets are priced differently depending on whether the future is perceived as a good or bad state of the nature. The model predicts an asymmetric effect of dividends on asset prices.

Given the non-linear nature of this result, co-integration techniques cannot be used to empirically test the model –as Campbell and Shiller (1987) did. Nevertheless, co-summability theory is precisely designed to test for long run relationship that combines at the same time persistence and non-linearity. This new theory is used to test the *threshold present value model* for the US stock market.

The chapter is organized as follows. In section 3.2, we describe the intertemporal decision problem of a threshold impatient investor and look for the optimal solution. To understand well this problem, we first put forward the decision problem of an standard investor. Then, the same problem for investors with threshold impatience is studied. Section 3.3 analyzes the long run implications for asset pricing of changing impatience. An important result of this section is that from the study of these long run properties an empirical model can be derived to test the theoretical one. Hence, we dedicate section 3.4 to explain the econometric techniques –co-summability theory– that make possible to test the derived model. In section 3.5, we use data on prices and dividends of the US stock market and the econometric techniques described in Section 3.4 to test for the *threshold present value model*. Section 3.6 finishes with some concluding remarks. Proofs are collected in the Appendix at the end of the chapter.

3.2 The Threshold Impatient Investor Problem

In this section we describe the threshold impatient investor problem and derive the conditions for an optimal solution, which we call the *threshold present value model*. For comparison, first, the problem of a standard investor is put forward.

3.2.1 Standard Investors

A standard investor solves

$$\max_{\{\xi_t\}} E_t \sum_{j=0}^{\infty} \beta^j u(c_{t+j}) \qquad s.t.$$

$$c_t = e_t - p_t \xi_t$$

$$c_{t+j} = e_{t+j} - p_{t+j}\xi_{t+j} + x_{t+j}\xi_{t+j-1} \qquad j = 1, 2, ...$$

where c_t and e_t denote consumption and endowment, respectively. ξ_t is the amount of asset the investor chooses to buy at price p_t and x_{t+j} is the *payoff* at time $t + j$. The period utility function $u(.)$ is assumed to be increasing and concave. Discounting the future by the subjective discount factor, β, captures impatience of the investor. Finally, E_t is the conditional expectation operator.

The first order condition –FOC– for the first two periods is

$$p_t = E_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right],$$

and, specifically, for stocks $x_{t+j} = p_{t+j} + d_{t+j}$, where d_{t+j} are dividends at time $t + j$. That is,

$$p_t = E_t \left[\beta \frac{u'(c_{t+1})}{u'(c_t)} (p_{t+1} + d_{t+1}) \right]. \tag{3.1}$$

Iterating equation (3.1) for $k = 1, 2, 3...$

$$p_{t+k} = E_{t+k} \left[\beta \frac{u'(c_{t+k+1})}{u'(c_{t+k})} (p_{t+k+1} + d_{t+k+1}) \right],$$

and imposing the transversality condition

$$\lim_{n\to\infty} E_t \left[\beta^n \frac{u'(c_{t+n})}{u'(c_t)} p_{t+n}\right] = 0,$$

the pricing formula is

$$p_t = E_t \sum_{j=1}^{\infty} \beta^j \frac{u'(c_{t+j})}{u'(c_t)} d_{t+j}. \tag{3.2}$$

When the investor is risk neutral or chooses a constant consumption path, equation (3.2) becomes

$$p_t = E_t \sum_{j=1}^{\infty} \beta^j d_{t+j}.$$

It is worth to emphasize that the same pricing formula is obtained from the definition of expected returns

$$E_t[r_{t+1}] = E_t \left[\frac{d_{t+1} + \Delta p_{t+1}}{p_t}\right]. \tag{3.3}$$

Equation (3.3) can be rewritten as

$$p_t = E_t \left[\frac{d_{t+1} + p_{t+1}}{1 + E_t[r_{t+1}]}\right]. \tag{3.4}$$

Iterating equation (3.4) for p_{t+1}, p_{t+2}, and so on, and imposing the transversality condition

$$\lim_{n\to\infty} E_t \left[\frac{p_{t+n}}{\prod\limits_{k=1}^{n} (1 + E_{t+j}[r_{t+j+1}])}\right] = 0,$$

the following solution is obtained

$$p_t = E_t \sum_{j=1}^{\infty} \frac{d_{t+j}}{\prod\limits_{k=1}^{j} (1 + E_{t+j}[r_{t+j+1}])}.$$

If $E_{t+j}[r_{t+j+1}] = r$, then

$$p_t = \sum_{j=1}^{\infty} \left(\frac{1}{1+r}\right)^j E_t[d_{t+j}],$$

which is the optimality condition of the standard investor problem in a risk neutral environment, with $\beta = 1/(1+r)$.

3.2.2 Threshold Impatient Investors

A threshold impatient investor solves

$$\max_{\{\xi_t\}} E_t \sum_{j=0}^{\infty} \beta(s_{t+j-1}) u(c_{t+j}) \qquad s.t.$$

$$c_t = e_t - p_t \xi_t$$

$$c_{t+j} = e_{t+j} - p_{t+j}\xi_{t+j} + x_{t+j}\xi_{t+j-1} \qquad j = 1, 2, ...$$

where, now, $\beta(s_{t+j-1})$ with $\beta(s_0) = 1$ is a time varying subjective discount factor, and s_{t+j-1} is the variable that determines the impatience of the investor. For instance, in Becker and Mulligan (1997) s_t is the spending resources on imagining future pleasures. However, instead of assuming that s_t is fully determined by the investor, we let the uncertainty in the economy to randomly affect the subjective discount factor via the threshold variable, s_t. Specifically, s_t will be considered a random process that captures quantitatively the perception of the future that the agents have. To be more precise, it will be assumed that $\beta(s_t) = \beta_1 1(s_t \geq \gamma) + \beta_2 1(s_t < \gamma)$, where $1(\cdot)$ is the indicator function. Therefore, the subjective discount factor will take two different values depending on whether the agent perceive the future as a good ($s_t \geq \gamma$) or bad ($s_t < \gamma$) state of the nature[11].

Proposition 12 *: In an economy with a threshold impatient representative investor the optimal asset pricing formula is*

$$p_t = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} \beta(s_{t+k-1}) \right) \frac{u'(c_{t+j})}{u'(c_t)} d_{t+j}.$$

Remark: If the investor is risk neutral or chooses a constant consumption path the pricing formula becomes

$$p_t = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} \beta(s_{t+k-1}) \right) d_{t+j}, \tag{3.5}$$

which we call the *threshold present value model.*

As in the standard investor case, the same present value relationship can be derived from the definition of expected returns assuming they follow a threshold process. Specifically, in that case

$$\beta(s_{t+k}) = \frac{1}{1 + r(s_{t+k})},$$

in equation (3.5) where $r(s_{t+k}) = E_{t+k}[r_{t+k+1}]$. Even in a risk-neutral environment, the C-CAPM can be reconciled with the time variation in expected returns.

3.3 Long run implications for asset pricing

Campbell and Shiller (1987) showed that when the subjective discount factor is constant and d_t is $I(1)$, then, by the standard present value formula, p_t and d_t are cointegrated with cointegrating vector $(1, -1/r)$. In other words, the model predicts the existence of a theoretical long run equilibrium relationship between prices and dividends. This result allows Campbell and Shiller (1987) to use cointegration analysis to test the present value model. The argument is as follows. Consider the

[11] Introducing more than two states of the nature is a straightforward generalization of the model. Nevertheless, to keep things simple, the theoretical analysis will be carried out considering only two states of the nature.

standard present value formula

$$p_t = E_t \sum_{j=1}^{\infty} \beta^j d_{t+j}, \tag{3.6}$$

where $\beta = 1/(1+r)$. Equation (3.6) can be rewritten as

$$p_t - \frac{1}{1+r} d_t - \left(\frac{1}{1+r}\right)^2 d_t - ... = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k E_t[(1-L^k) d_{t+k}],$$

or equivalently

$$p_t - \frac{1}{r} d_t = \sum_{k=1}^{\infty} \left(\frac{1}{1+r}\right)^k E_t[(1-L^k) d_{t+k}]. \tag{3.7}$$

Therefore, if d_t is $I(1)$, then prices and dividends are cointegrated with cointegrating parameter $1/r$.

Following Campbell and Shiller (1987), next proposition shows that if the subjective discount factor follows a threshold process and agents are risk neutral, then prices and dividends should still share a long run equilibrium relationship. However, in this case the equilibrium is of a threshold type.

Assumption 1:

(i) $d_t = d_{t-1} + u_t$, $d_0 = 0$

(ii) $\beta(s_t) = \beta_1 \mathbf{1}(s_t \geq \gamma) + \beta_2 \mathbf{1}(s_t < \gamma)$, $s_t \sim i.i.d.$

(iii) at time t, $p(s_{t+j} \leq \gamma | I_t) = p$, $j = 1, 2, ...$

(iv) u_s, and s_r are $i.i.d.(0,1)$ and uncorrelated for every s and r.

Proposition 13 *: Under Assumptions 1, if*

$$p_t = E_t \sum_{j=1}^{\infty} \left(\prod_{k=1}^{j} \beta(s_{t+k-1}) \right) d_{t+j},$$

then

$$p_t - \phi_1 1(s_t \geq \gamma) d_t - \phi_2 1(s_t < \gamma) d_t = \sum_{j=1}^{\infty} \beta(s_t) \pi^j E_t[(1-L^j) d_{t+j}], \tag{3.8}$$

where $\pi = \beta_1 p + \beta_2 (1-p)$, $\phi_1 = \beta_1 / (1-\pi)$, $\phi_2 = \beta_2 / (1-\pi)$.

Remark: Equation (3.8) rewritten as

$$p_t - \phi_1 1(s_t \geq \gamma) d_t - \phi_2 1(s_t < \gamma) d_t = v_t, \tag{3.9}$$

offers an empirical framework in a regression model form to test the threshold present value model. Notice, however, that equation (3.9) involves non-linear transformations of a persistent processes $-1(s_t \geq \gamma) d_t$ and $1(s_t < \gamma) d_t$. In other word, a comparison between equations (3.7) and (3.9) shows that while in the former equation the co-integrating vector is constant, in the latter it is of a threshold type.

3.4 Co-summability Theory

The ideas of integration and co-integration cannot be directly used to analyze non-linear equilibrium relationships among persistent variables since these concepts do not properly apply. Consider the following non-linear relationship $y_t = \theta f(x_t) + u_t$. If it were known that $f(x_t)$ is $I(d)$, then the standard framework of co-integration would fit perfectly. However, when x_t is persistent, say $I(1)$, then for many interesting non-linear transformations f the order of integration of $f(x_t)$ will not be well defined. It will be convenient to consider the following case.

Example 13 *: Product of Indicator Function and Random Walk*

Let

$$h_t = 1(s_t \leq \gamma) d_t. \tag{3.10}$$

The variance and autocovariances of h_t depend on time, hence, one would think that this process is $I(1)$. However, the variance of the first difference

$$V[\Delta h_t] = V[1(s_t \leq \gamma)d_t - 1(s_{t-1} \leq \gamma)d_{t-1}] = [2p(1-p)]t + p(2p-1).$$

In fact, it can be considered that $h_t \sim I(\infty)$, in the sense that, the variance of $\Delta^d h_t$ depends on t regardless of the value of d –see Yoon (2005). However, this characterization is not useful neither to characterize univariate processes nor generalize co-integration theory to non-linear worlds.

In Chapters 1 and 2 co-summability theory –a generalization of co-integration to non-linear long run equilibrium relationships– is developed. Co-summability is built upon the concept of order of summability of stochastic processes. It was first introduced in a heuristic way in Gonzalo and Pitarakis (2006) and subsequently formalized in Chapter 1 of this thesis.

Definition 8 *: A stochastic process y_t with positive variance is said to be summable of order δ, represented as $S(\delta)$, if*

$$S_n = \frac{1}{n^{\frac{1}{2}+\delta}} L(n) \sum_{t=1}^{n} (y_t - m_t) = O_p(1) \qquad \text{as } n \to \infty,$$

where δ is the minimum real number that makes S_n bounded in probability, m_t is a deterministic sequence, and $L(n)$ is a slowly-varying function[12]*.*

[12] A positive, Lebesgue measurable function L, on $(0, \infty)$ is slowly varying –in the Karatama's sense– at ∞ if

$$\frac{L(\lambda n)}{L(n)} \to 1 \ \ (n \to \infty) \ \forall \lambda > 0.$$

(See Embrechts, Klüppelberg and Mikosh, 1999, p.564).

The order of summability, δ, gives a summary measure of the stochastic properties –persistence and evolution of the variance– of y_t without relying on a particular data generating process. In this sense, it can overcome the drawbacks of using the order of integration in non-linear environments.

Notice that, when possible, the order of summability will be determined by some Central Limit result. In particular, in the case of h_t the following characterization is found.

Summability in Example 13 (*Product of indicator function and random walk*): Applying the Functional Centra Limit Theorem –FCLT– and the Continuous Mapping Theorem –CMT–,

$$S_n = \frac{1}{n^{\frac{3}{2}} p} \sum_{t=1}^{n} 1(s_t \leq \gamma) d_t \Longrightarrow \int_0^1 W(r)dr,$$

implying that $1(s_t \leq \gamma)d_t$ is $S(1)$ with, for instance, $L(n) = 1/p$.

In the same way integration constitutes the first step to check the balancedness of a linear relationship and to analyze co-integration, summability can be used to study non-linear long run relationships.

Definition 9 *: A postulated relationship*

$$y_t = f(x_t, \theta),$$

will be said to be balanced if $y_t \sim S(\delta_y)$, $z_t = f(x_t, \theta) \sim S(\delta_z)$, *and* $\delta_y = \delta_z$.

Once the balancedness of a non-linear model is established, the analysis of non-linear long run relationships can be done using the concept of co-summability.

Definition 10 *: Two summable stochastic processes,* $y_t \sim S(\delta_y)$ *and* $x_t \sim S(\delta_x)$, *will be said to be co-summable if there exists* $z_t = f(x_t, \theta) \sim S(\delta_y)$ *such that* $u_t = y_t - f(x_t, \theta)$ *is* $S(\delta_u)$, *with* $\delta_u = \delta_y - \delta$ *and* $\delta > 0$. *In short,* $(y_t, z_t) \sim CS(\delta_y, \delta)$.

Econometric tools developed in Chapters 1 and 2 allow to (i) empirically estimate and infer the order of summability of an observed time series and (ii) test for balancedness and co-summability.

Let the least squares regression

$$y_t = \hat{\theta} f(x_t) + \hat{u}_t,$$

where $f : \mathbb{R} \to \mathbb{R}$, x_t, and y_t are known by the researcher. Let $z_t = f(x_t)$. The following assumption ensures that y_t and z_t are $S(\delta_y)$ and $S(\delta_z)$, respectively.

Assumption 2.

$$S_{yn} = \frac{1}{n^{1/2+\delta_y}} \sum_{t=1}^{n} y_t = O_p(1) \quad \text{and} \quad S_{zn} = \frac{1}{n^{1/2+\delta_z}} \sum_{t=1}^{n} f(x_t) = O_p(1).$$

Following exactly the same logic of co-integration theory, the following empirical strategy is devised.

Step 1: Test for Balancedness

Step 2: Test for Strong Co-summability

The test for balancedness uses the order of summability estimator developed in Chapter 1. It follows the convergence rate estimation procedure in McElroy and Politis (2007), which is based on a simple least squares regression. The procedure requires the following assumption.

Assumption 3. $P(S_n = 0) = 0$ for all $n = 1, 2, 3, ...$

Under Assumptions 2 and 3,

$$U_{yk} = \log S_{yk}^2 = \log \left[\left(\frac{1}{k^{\frac{1}{2}+\delta_y}} \sum_{t=1}^{k} y_t \right)^2 \right] = O_p(1),$$

and the following regression model can be derived

$$Y_{yk} = \beta_y \log k + U_{yk}, \tag{3.11}$$

where $Y_{yk} = \log \left(\sum_{t=1}^{k} y_t \right)^2$ and $\beta_y = 1 + 2\delta_y$.

Chapter 1 shows that the ordinary least squares –OLS– estimator of $\beta_y = 1 + 2\delta_y$ is $\log n$-consistent with an asymptotic distribution that cannot be tabulated in general. Through simulations, it is shown that subsampling confidence intervals can be constructed to undertake inferences on the true δ_y. It is important to mention that the presence of deterministic components in the DGP has a strong bias effect on the order of summability estimator, at least in finite samples. In Chapter 1 valid demeaning and detrending procedures are developed. Nevertheless, to facilitate exposition no deterministic components will be considered in this section.

Notice that the regression to estimate the order of summability of z_t is

$$Y_{zk} = \beta_z \log k + U_{zk}, \tag{3.12}$$

where $Y_{zk} = \log \left(\sum_{t=1}^{k} z_t \right)^2$ and $\beta_z = 1 + 2\delta_z$.

To test for balancedness, an auxiliary equation that substracts (3.12) from (3.11) will be used, that is,

$$Y_{yk} - Y_{zk} = \left(\beta_y - \beta_z \right) \log k + U_{yk} - U_{zk}.$$

Let $Y_k = Y_{yk} - Y_{zk}$, $\beta = \beta_y - \beta_z$, and $U_k = U_{yk} - U_{zk}$. Then, testing $H_o : \delta_y = \delta_z$ is equivalent to test $H_o : \beta = 0$ in

$$Y_k = \beta \log k + U_k. \tag{3.13}$$

It is shown in Chapter 2 that the OLS estimator $\hat{\beta}_n$ of β is a consistent estimator of the difference $\beta_y - \beta_z$. In particular, under balancedness

$$\hat{\beta}_n \xrightarrow{p} 0.$$

Its asymptotic distribution cannot be tabulated in general. Nonetheless, subsampling confidence intervals allow to undertake inference.

Balancedness is a necessary, although not sufficient, condition for having correctly specified a model. In fact, a relationship could be balanced spuriously. To distinguish between spurious and co-summable relations a test for co-summability is next described to implement Step 2 of the empirical strategy. Basically, co-summability is tested through a residual based statistic. The procedure consists on estimating and inferring if the order of summability of the OLS residuals[13] is zero – $H_o : \delta_{\hat{u}} = 0$. To this end, the order of summability estimator and subsampling confidence intervals are used.

Remark: Let

$$y_t = m + \theta f(x_t) + u_t,$$

where m is an unknown constant term. In addition, let $\hat{m}$ and $\hat{\theta}$ the OLS estimator of m and θ, respectively. In this case,

$$\sum_{t=1}^{n} \hat{u}_t = 0,$$

which implies that $\hat{u}_t$ cannot be used to infer δ_u. The following pseudo residuals

$$\tilde{u}_t = y_t - \hat{\theta} f(x_t) = m + u_t - (\hat{\theta} - \theta) f(x_t),$$

could be used instead since

$$\sum_{t=1}^{n} \tilde{u}_t \neq 0.$$

The effect of m on the estimation of the order of summability of u_t can be tackled by an appropriate partial demeaning procedure –see Section 1.4.4 in Chapter 1.

3.5 Empirical Results

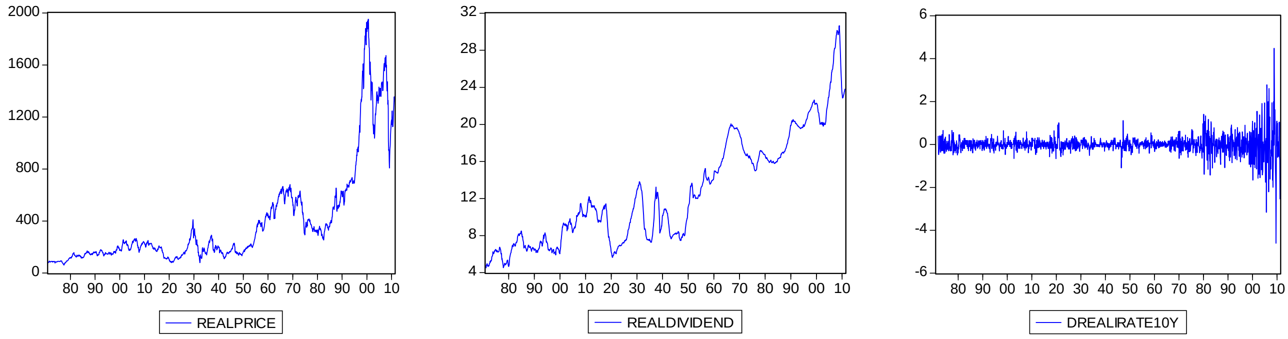
In this section we apply co-summability theory to test the *threshold present value model*. To this end, we use US monthly time series covering the period 1872:01-2011:03, which are obtained from

[13] It is worth to mention that when unbalanced or spurious relationships are considered there is not a properly defined error in the model since, in fact, there is not a relationship in those cases. Moreover, the order of summability of the residuals equals the order of summability of the errors under co-summability and the order of summability of the endogenous variable in the model under spurious and unbalanced situations. Therefore, a test for co-summability can be constructed by testing the order of summability of the residuals –see Section 2.3. in Chapter 2 for specific details.

Robert J. Shiller. Specifically, we use (i) real S&P stock price index, (ii) real dividends, and (iii) changes of ten years real government yield as a proxy of the threshold variable. Long term interest rates should take into account the expectations of the agents and hence their perception of the future. Therefore, a decrease in long term interest rate should reflect expectations of bad times to come, while an increase should reflect agents expectations of an economic boom.

Table 20: C-CAPM with Threshold Impatient Investors

Variables



The estimated orders of summability of these variables, –p_t, d_t, and Δi_t– and its corresponding subsampling confidence intervals are reported in Table 21. It is important to mention that to control for a possible constant term in regression model (3.11) the first observation is substracted. The corresponding order of summability estimator is denoted by $\hat{\delta}^*$. Moreover, to deal with the effect of the deterministic components in these time series, proper detrending procedures has been carried out. Specifically, Δi_t has been partially demeaned. For the other variables, a partial quadratic detrending procedure has been applied –see Section 1.4.4. in Chapter 1 for details. Attending to the results in Table 21, it can be seen that while the order of summability of p_t and d_t are 1.3 and 1, respectively, the threshold variable, Δi_t, appers to be summable of order 0.1. Notice that one belongs to the confidence intervals of both p_t and d_t but not to that of Δi_t.

Next, the following two specifications will be analyzied. First, a linear relationship *à la Campbell and Shiller*

$$p_t = \theta_0 d_t + u_{0t},$$

Table 21: Order of Summability: Estimation and Inference

Variables	$\hat{\delta}^*$	I_L	I_U
p_t	1.335	0.697	1.974
d_t	1.046	0.547	1.546
Δi_t	0.105	-0.397	0.608

$\hat{\delta}^*$ denotes the estimated order of summability from a regression that substracts the first observation. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals.

and second, a threshold model

$$p_t = \theta_1 d_t 1\left(\Delta i_t \geq \gamma\right) + \theta_2 d_t 1\left(\Delta i_t < \gamma\right) + u_t,$$

to test the threshold present value estatment. For the threshold specification, several values of the threshold parameter, $\gamma = \{0, 1, 2, 3, 4\}$, are considered. All the specifications were found to be balanced; hence, Table 22 presents only the results of testing for strong co-summability.

Table 22: Testing for Co-summability

OLS	p_t	p_t	p_t	p_t	p_t	p_t
		$\gamma = 0$	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$
d_t	35.794					
$d_t 1\left(\Delta i_t \geq \gamma\right)$		35.306	21.889	20.757	19.100	19.232
$d_t 1\left(\Delta i_t < \gamma\right)$		36.245	36.796	36.173	35.951	35.880
$\hat{\delta}^*_{\hat{u}}$	0.873	0.825	0.359	0.347	0.318	0.883
I_L	0.408	0.379	-0.023	-0.002	-0.041	0.400
I_U	1.338	1.270	0.741	0.696	0.677	1.367

$\hat{\delta}_u$ denote the estimated order of summability of the residuals in each regression. I_L and I_U denote the lower and upper bounds of the corresponding subsampling intervals.

The estimated order of the residuals, $\hat{\delta}^*_{\hat{u}}$, in the linear model –column 2 of Table 22– is around 0.8 and this value belongs to the subsampling confidence interval associated to the order of summability of p_t in Table 21. Moreover, the corresponding subsampling confidence interval indicates that the model does not specifies a co-summable relationship. With respect the threshold specifications, the results depend on the value of the threshold parameter. When $\gamma = 0$ or $\gamma = 4$, the model does not specify a co-summable relationship. Nonetheless, when $\gamma = 1$, $\gamma = 2$, or $\gamma = 3$, a co-summable relationship is found. In any case, all these co-summable cases point to an asymetric relationship between prices and dividens. Attending to the point estimates, the reaction of prices to dividends appears to be lower when investors expect a boom than otherwise. Hence, this result point to a level of patience of the investors that is higher when bad times are expected to come in the future –a result that is in line with a stream of the experimental literature finding that people are more patience for looses than for gains.

3.6 Conclusions

In an economy with threshold impatient investors, assets are priced via a *threshold present value model.* Under such a pricing formula, an asymmetric long run equilibrium relationship between prices and dividends is derived. Using recent co-summability techniques, the model is tested empirically using US stock market data. Results point to a non-linear relationship between prices and dividends. Nevertheless, these empirical results demand a deeper and more detailed analysis. For instance, other types of threshold variables could be taken as a proxy of expectations in the economy or the threshold parameter should be freely estimated.

From a more general and theoretical perspective, this chapter opens a full project for future research. Given the theoretical threshold relationship between prices and dividends, a multivariate analysis *à la Campbell and Shiller (1987)* can be done. Specifically, the restrictions in an error correction model should be derived in order to test the *threshold present value model* in a multivariate setup. This implies to undertake a generalization of threshold vector error correction models –TVECM– with its corresponding econometric treatment. Other potential threshold economic relationships could, then, be tested using this general TVECM.

3.7 Appendix

Proof of Proposition 12: The FOC for the two first periods is

$$p_t = E_t \left[\beta(s_t) \frac{u'(c_{t+1})}{u'(c_t)} x_{t+1} \right],$$

and specifically for stocks it is

$$p_t = E_t \left[\beta(s_t) \frac{u'(c_{t+1})}{u'(c_t)} (p_{t+1} + d_{t+1}) \right].$$

Now we can iterate for

$$p_{t+1} = E_{t+1} \left[\beta(s_{t+1}) \frac{u'(c_{t+2})}{u'(c_{t+1})} (p_{t+2} + d_{t+2}) \right],$$

in equation (3.1) to get

$$\begin{aligned} p_t &= E_t\left[\beta(s_t)\frac{u'(c_{t+1})}{u'(c_t)}d_{t+1}\right] + E_t\left[\beta(s_t)\beta(s_{t+1})\frac{u'(c_{t+2})}{u'(c_t)}d_{t+2}\right] \\ &\quad + E_t\left[E_{t+1}\left[\beta(s_t)\beta(s_{t+1})\frac{u'(c_{t+2})}{u'(c_t)}p_{t+2}\right]\right]. \end{aligned}$$

Since

$$p_{t+2} = E_{t+2}\left[\beta(s_{t+2})\frac{u'(c_{t+3})}{u'(c_{t+2})}\left(p_{t+3}+d_{t+3}\right)\right],$$

we have

$$\begin{aligned} p_t &= E_t\left[\beta(s_t)\frac{u'(c_{t+1})}{u'(c_t)}d_{t+1}\right] + E_t\left[\beta(s_t)\beta(s_{t+1})\frac{u'(c_{t+2})}{u'(c_t)}d_{t+2}\right] \\ &\quad + E_t\left[\beta(s_t)\beta(s_{t+1})\beta(s_{t+2})\frac{u'(c_{t+3})}{u'(c_t)}d_{t+3}\right] \\ &\quad + E_t\left[E_{t+1}\left[E_{t+2}\left[\beta(s_t)\beta(s_{t+1})\beta(s_{t+2})\frac{u'(c_{t+3})}{u'(c_t)}p_{t+3}\right]\right]\right]. \end{aligned}$$

Iterating forward and imposing the transversality condition

$$\lim_{n\to\infty} E_t\left[\beta(s_t)\cdot\ldots\cdot\beta(s_{t+n-1})\frac{u'(c_{t+n})}{u'(c_t)}p_{t+n}\right] = 0,$$

the final FOC is

$$p_t = E_t\sum_{j=1}^{\infty}\left(\prod_{k=1}^{j}\beta(s_{t+k-1})\right)\frac{u'(c_{t+j})}{u'(c_t)}d_{t+j}.$$

Q.E.D.

Proof of Proposition 12: We have shown that if investors are risk neutral but their impatience changes along time

$$p_t = E_t\sum_{j=1}^{\infty}\left(\prod_{k=1}^{j}\beta(s_{t+k-1})\right)d_{t+j}. \tag{3.14}$$

Given assumption (v) we can rewrite equation (3.14) as

$$p_t = \sum_{j=1}^{\infty}E_t\left[\prod_{k=1}^{j}\beta(s_{t+k-1})\right]E_t\left[d_{t+j}\right].$$

By assumptions (iii) and (iv)

$$\begin{aligned} E_t\left[\left(\prod_{k=1}^{j}\beta(s_{t+k-1})\right)\right] &= \beta(s_t)\prod_{k=2}^{j}E_t\left[\beta(s_{t+k-1})\right] \\ &= \beta(s_t)\left(\beta_1 p + \beta_2(1-p)\right)^{j-1}. \end{aligned}$$

And hence, the price of the stock is

$$\begin{aligned} p_t &= \sum_{j=1}^{\infty} \beta(s_t)\left(\beta_1 p + \beta_2(1-p)\right)^{j-1} E_t[d_{t+j}] \\ &= \beta(s_t)\sum_{j=1}^{\infty}\left(\beta_1 p + \beta_2(1-p)\right)^{j-1} E_t[d_{t+j}] \\ &= \beta(s_t)\sum_{j=1}^{\infty}\pi^{j-1} E_t[d_{t+j}], \end{aligned}$$

where $\pi = \beta_1 p + \beta_2(1-p)$. Note now that

$$p_t - \beta(s_t)d_t - \beta(s_t)\pi d_t - \beta(s_t)\pi^2 d_t - ... = \sum_{j=1}^{\infty}\beta(s_t)\pi^j E_t[(1-L^j)d_{t+j}].$$

Hence,

$$p_t - \beta(s_t)\frac{1}{1-\pi}d_t = \sum_{j=1}^{\infty}\beta(s_t)\pi^j E_t[(1-L^j)d_{t+j}]$$

or equivalently,

$$p_t - \left[\frac{\beta_1}{1-\pi}\mathbf{1}(s_t \geq \gamma) + \frac{\beta_2}{1-\pi}\mathbf{1}(s_t < \gamma)\right] d_t = \sum_{j=1}^{\infty}\beta(s_t)\pi^j E_t[(1-L^j)d_{t+j}].$$

The right hand side of equation (3.8) is stationary, hence it is the case for the threshold combination relating prices and dividends, the left hand side. Therefore, these variables should be cointegrated with threshold effects in the long run relationship.

Q.E.D.

References

Ainslie, G. (1992): *Picoeconomics: The strategic interaction of successive motivational states within the person*', New York: Cambridge University Press.

Ainslie, G. and Haslam, N. (1992): "Hyperbolic discounting," in *Choice over Time*, Loewenstein, G. and J. Elster (editors), Russell Sage Foundation.

Arteche, J., and J. Orbe (2005): "Bootstrapping the log-periodogram regression," *Economics Letters*, 86, 79-85.

Becker, G. S. and Mulligan, C.B. (1997): "The endogenous determination of time preference," *The Quarterly Journal of Economics*, 112, 729-758.

Beran, J. (1994): *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

Brock, W. A., and M. S. Taylor (2005): "Economic growth and the environment: A review of theory and empirics," *Handbook of Economic Growth*, Volume 1, Part 2, 1749-1821.

Campbell, J. Y and Shiller, R. J. (1987): "Cointegration and tests of present value models," *Journal of Political Economy*, 95, 1062-88.

Choi, I., and P. Saikkonen (2010): "Tests for nonlinear cointegration," *Econometric Theory*, 26, 682-709.

Christensen, A. M., and H. B. Nielsen (2009): "Monetary Policy in the Greenspan Era: A Time Series Analysis of Rules vs. Discretion," *Oxford Bulletin of Economics and Statistics*, 71, 69-89.

Clarida, R., and M. Gertler (1997): "How the Bundesbank conducts monetary policy". In: Romer, D. (Ed.), *Reducing Inflation*. Chicago: University of Chicago Press.

Clarida, R., J. Galí, and M. Gertler (1998): "Monetary policy rules in practice. Some international evidence," *European Economic Review*, 42, 1033-1067.

Clarida, R., J. Galí, and M. Gertler (2000): "Monetary policy rules and macroeconomic stability: Evidence and some theory," *Quarterly Journal of Economics*, 115, 147-180.

Dasgupta, S., B. Laplante, H. Wang, and D. Wheeler (2002): "Confronting the Environmental Kuznets Curve," *The Journal of Economic Perspectives*, 16, 147-168.

Davidson, J. (2009): "When is a time series I(0)?," *The methodology and practice of econometrics*, a festschrift for David F. Hendry edited by Jennifer Castle and Neil Shepherd, Oxford University Press.

de Jong, R. M., and Ch. H. Wang (2005): "Further results on the asymptotics for non-linear transformations of integrated time series," *Econometric Theory*, 21, 413-430.

Dickey, D.A., and W. A. Fuller (1979): "Distribution of the estimator for autoregressive time series with a unit root," *Journal of the American Statistical Association*, 74, 427–431.

Dolado, J. J., R. María-Dolores, and M. Naveira (2005): "Are monetary-policy reaction functions asymmetric?: The role of nonlinearity in the Phillips curve," *European Economic Review*, 49, 485-503.

Embrechts, P., C. Klüppelberg, and T. Mikosh (1999): *Modelling extremal events for insurance and finance.* Berlin: Springer-Verlag.

Engle, R.F., and C. W. J. Granger (1987): "Co-integration and error correction: representation, estimation and testing," *Econometrica*, 55, 251–276.

Fan, J., and Q. Yao, (2003): *Nonlinear time series: Nonparametric and parametric methods.* New York: Springer-Verlag.

Franses, P. H., and D. van Dijk (2000): *Non-linear time series models in empirical finance.* Cambridge: Cambridge University Press.

Geweke, J. and S. Porter- Hudak (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4, 221-238.

Giraitis, L., P. M. Robinson, and D. Surgailis (1999): "Variance-type estimation of long memory," *Stochastic Processes and their Applications, 80, 1-24.*

Gonzalo, J., and J. Y. Pitarakis (2006): "Threshold effects in co-integrating regressions," *Oxford Bulletin of Economics & Statistics*, 68, 813-833.

Granger, C. W. J. (1995): "Modelling non-linear relationships between extended-memory variables," *Econometrica*, 63, 265-279.

Granger, C. W. J., and J. Hallman (1991): "Non-linear transformations of integrated time series," *Journal of Time Series Analysis*, 12, 207-224.

Granger, C. W. J., and T. Teräsvirta (1993): *Modelling non-linear economic relationships.* Oxford: Oxford University Press.

Grossman, G. M., and A. B. Krueger (1995): "Economic growth and environment," *Quarterly Journal of Economics*, 110, 353-377.

Hansen, B. E. (1992): "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory*, 8, 489-500.

Holtz-Eakin, D., and T. M. Selden (1995): "Stoking the fires? CO_2 emissions and economic growth," *Journal of Public Economics*, 57, 85-101.

Hong, S. H., and M. Wagner (2008): "Nonlinear cointegration analysis and the environmental Kuznets curve," Mimeo.

Hurst, H. E. (1951): "Long-term storage capacity of reservoirs," *Transactions of the American Society of Civil Engineers*, 116, 770-808.

Jalil, A., and S. F. Mahmud (2009): "Environment Kuznets curve for CO_2 emissions: A cointegration analysis for China," *Energy Policy*, 37, 5167-5172.

Johansen, S. (1991): "Estimation and hypothesis testing of co-integration vectors in Gaussian vector autoregressive models," *Econometrica*, 59, 1551-1580.

Johansen, S. (1995): *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford: Oxford University Press.

Jones, Ch. I. (1995): "Time series tests of endogenous growth models," *Quarterly Journal of Economics*, 110, 495-525.

Karlsen, H. A., T. Myklebust, and D. Tjϕstheim (2007): "Nonparametric estimation in a nonlinear cointegration type model," *The Annals of Statistics*, 35, 252-299.

Leybourne, S. J., B. P. M. McCabe, and A. R. Tremayne (1996): "Can economic time series be differenced to stationarity?" *Journal of Business and Economic Statistics*, 14, 435-446.

Liu, M. (1998): "Asymptotics of non-stationary fractional integrated series," *Econometric Theory*, 14, 641-662.

Loewenstein, G. and Prelec, D. (1992): "Anomalies in intertemporal choice: evidence and an interpretation," *The Quarterly Journal of Economics*, 107, 573–597.

Lucas, R. E. (1978): "Asset prices in an exchange economy," *Econometrica*, 46, 1429-1446.

McElroy, T., and D. N. Politis (2007): "Computer-intensive rate estimation, diverging statistics, and scanning," *The Annals of Statistics*, 35, 1827-1848.

Müller, U. K. (2008): "The impossibility of consistent discrimination between $I(0)$ and $I(1)$ processes," *Econometric Theory*, 24, 616-630.

Nelson, Ch. R., and Ch. I. Plosser (1982): "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139-162.

Österholm, P. (2005), "The Taylor Rule: A Spurious Regression?," *Bulletin of Economic Research*, 57, 217-247.

Park, J. Y., and P. C. B. Phillips (1988): "Statistical inference in regressions with integrated processes: Part 1," *Econometric Theory*, 4, 468–497.

Park, J. Y., and P. C. B. Phillips (1999): "Asymptotics for non-linear transformations of integrated time series," *Econometric Theory*, 15, 269-298.

Park, J. Y., and P. C. B. Phillips (2001): "Nonlinear regressions with integrated time series," *Econometrica*, 69, 117-161.

Perman, R. J., and D. Stern (2003): "Evidence from panel unit root and cointegration tests that the Environmental Kuznets Curve does not exist," Australian Journal of Agricultural and Resource Economics, 47, 325-347.

Phillips, P. C. B. (1986): "Understanding spurious regressions in econometrics," *Journal of*

Econometrics, 33, 311-340.

Phillips, P. C. B., and V. Solo (1992): "Asymptotics for linear processes," *Annals of Statistics*, 20, 971–1001.

Politis, D. N., J. P. Romano, and M. Wolf (1999): *Subsampling.* New York: Springer.

Pötscher, B. M. (2004): "Non-linear functions and convergence to Brownian motion: beyond the continuous mapping theorem," *Econometric Theory*, 20, 1-22.

Romano, J. P., and A. F. Siegel (1986): *Counterexamples in probability and statistics.* Monterey, California: Wadsworth and Brooks/Cole.

Schienle, M. (2011): "Nonparametric nonstationary regression with many covariates," Mimeo.

Siklos, P. L., and M. E. Wohar (2005): "Estimating Taylor-type rules: An unbalanced regression?," *Econometric Analysis of Financial and Economic Time Series*, Volume 20, Part B, 239-276.

Stock, J. H. (1991): "Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series," *Journal of Monetary Economics*, 28, 435-459.

Svensoon, L. E. O. (1997): "Inflation forecast targeting: Implementing and monitoring inflation targets," *European Economic Review*, 41, 1111-1146.

Taylor, J. B. (1993): "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Teräsvirta, T., D. Tjϕstheim, and C. W. J. Granger (2011): *Modelling nonlinear economic time series.* Forthcoming.

White, J. S. (1958): "The limiting distribution of the serial correlation coefficient in the explosive case," *Annals of Mathematical Statistics*, 29, 1188-1197.

Wang, Q., and P. C. B. Phillips (2009): "Structural nonparametric cointegrating regressions," *Econometrica*, 77, 1901-1948.

Yoon, G. (2005): "An introduction to $I(\infty)$ processes," *Economic Modelling*, 22, 473-483.